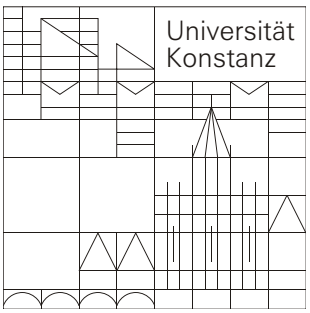
Visuelle Analyse von RFID-Sensordaten mit Raum- und Zeitbezug zur Untersuchung von Mausbewegungen

Bachelor-Arbeit an der Universität Konstanz im Fachbereich Informatik und Informationswissenschaft

vorgelegt von

Halldór Janetzko



September 2008

Einreichung: 2. September 2008

Prüfer: Prof. Dr. Daniel A. Keim, Universität Konstanz

Prof. Dr. Oliver Deussen, Universität Konstanz

Betreuer: Dr. Florian Mansmann

Zusammenfassung

Positionsdaten von Käfigmäusen wurden mittels RFID - Technologie erfasst, um Verhaltensweisen von gesunden und mit Alzheimer infizierten Mäusen zu unterscheiden. Die millionengroße und millisekundengenaue Datenmenge wurde mittels verschiedenen Visualierungstechniken analysiert. Allein bei der Analyse von Laufdistanzen oder Trinkhäufigkeiten konnten schon Unterschiede zwischen den Mäusen festgestellt werden. Durchschnittliche Tagesabläufe der Mäuse wurden zum Clustern verwendet und das Clustern konnte einen Teil der Mäuse richtig gruppieren. Die gleichzeitige Berücksichtigung von zeitlicher und räumlichen Dimensionen führte zur besten Unterscheidung der Daten. Mittels einer Aufteilung der Lebensmonate einer jeden Maus konnte eine lebensumfassende Visualisierung der Mäuse geschaffen werden. Die beste Unterscheidung unter Berücksichtigung des räumlichen und zeitlichen Bezugs der Daten gelang bezüglich des Geschlechtes der Mäuse.

Abstract

Position data of in cage living mice by using RFID technology were collected in order to distinguish behaviour of healthy and ill (infected with Alzheimer's disease) animals. Due to the extensive amount of per-millisecond log entries visualization is used to analyse the log data. Already by visualizing distances run by mice differences between ill and healthy or female and male respectively can be found. Furthermore visits of water places are examined to find changes in drinking habits. Additionally average daily routines were used for clustering and resulted in a correct partition of parts of the data. The best classification of all mice could be found by using temporal and geospatial dimensions at the same time. Lifelong visualizations of mice were generated by partitioning the lifetime and combining them within one image. All mice could be separated best by their gender, because the differences between the genders were much bigger than the differences between the ill and healthy mice.

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einführung

In vielen Lebensbereichen entstehen räumliche und zeitabhängige Daten, seien es unsere täglichen Bewegungen (zu jedem Zeitpunkt ist man an einem bestimmten Ort) oder beispielsweise GPS-Daten in der Logistikbranche zur besseren Ausnutzung der Transportkapazitäten. Häufig sind es sehr große Mengen millisekundengenauer Daten mit einer sehr hohen Auflösung. Dabei ist eigentlich nicht nur wichtig wann und wo sich ein Objekt aufhielt, sondern auch welche anderen Objekte in Interaktion mit diesem Objekt getreten sind oder welche Abhängigkeiten bzw. Beziehungen zwischen den Objekten bestanden. Unter Umständen soll ein Profil eines oder aller Objekte erstellt werden und die Abweichungen vom Profil aufgezeigt werden. Heutzutage ist es möglich beliebige Bewegungen aufzuzeichnen, beispielsweise von Menschen mittels Kameraüberwachung oder Kreditkartennutzung, von Lastwagen mittels Mautabrechnungen oder von Tieren durch GPS - Peilsender.

Die Erkenntnisse, die aus diesen Daten gezogen werden, können auf viele Arten genutzt werden. Denkt man an den „Problembären Bruno“, der im Sommer 2006 in Bayern erlegt wurde, hätte vielleicht eine Analyse des Bewegungsprofils den Bären retten können. Wäre schon früh aufgefallen, dass der Bär sich am liebsten in der Nähe von menschlichen Siedlungen aufhält und auch gerne Schafherden angreift, hätte man ihn beispielsweise in einen Zoo oder in eine menschenleere Gegend bringen können. Aber selbst wenn damals die Bewegungen des Bären genau aufgezeichnet worden wären, hätte immer noch ein automatisiertes oder halbautomatisiertes Tool für eine Analyse verwendet werden müssen, da ein Mensch kaum in der Lage gewesen wäre, alle Daten von Hand zu analysieren. Mit einer guten Visualisierungs- und Analysetechnik könnte man also viel erreichen, sei es beim Tierschutz, in der Wirtschaft oder in der Verbrechensbekämpfung. Die einzigen Probleme liegen – von moralischer Bedenken abgesehen – in der Auswertung und Analyse dieser (oft sehr großen) Datenmengen.

Es nutzt dem Informationsanalysten nichts, wenn er sich beispielsweise die Bewegungen aller Lastwagen als Animation beliebig oft anschauen kann, da das menschliche Gehirn nicht darauf ausgelegt ist, Änderungen zu bemerken. Die Versuche zum Thema „change blindness“ beweisen eindrücklich, wie wenig das Gehirn merkt, auch wenn es weiß, dass es nach Änderungen Ausschau halten soll (vgl. Kapitel 1.3 „The human user“ in [15]). Die Auswertung und Analyse dieser Daten ist also eine große Herausforderung und noch immer nicht vollständig gelöst. Die größte Schwierigkeit dabei ist die Verbindung von Zeit und Raum: räumliche Komponenten können beispielsweise durch Karten wiedergegeben werden und zeitliche Komponenten beispielsweise durch Zeitreihen. Nur die Verknüpfung dieser Dimensionen impliziert eine Art Videowiedergabe des Geschehens. Diese ist natürlich nicht sinnvoll, um eine Datenanalyse effizient und effektiv durchzuführen. Sinnvoll ist es, wenn der Nutzer auf einen Blick die wichtigsten Zeitpunkte erkennen und analysieren kann. Doch genau hier treten die Probleme auf: es ist sehr schwierig, algorithmisch die interessanten Zeitpunkte festzustellen. Wenn der Computer allerdings die geistigen Fähigkeiten des Menschen ausnützt, indem er die

Daten in einer geeigneten Visualisierung präsentiert und der Mensch die für ihn interessanten Zeitpunkte betrachten kann, ist die Lösung des Problems näher gerückt. Die Wahl der geeigneten Visualisierung ist der wichtigste Teil, um dem Benutzer ein gutes Werkzeug an die Hand zu geben.

Die Daten, die im Rahmen meines Bachelorprojektes untersucht wurden, stammen aus einem Experiment zur Verhaltensforschung an Mäusen in der Universität von Münster [13]. Dort wurde mit genetisch veränderten Mäusen experimentiert, welche nun die Veranlagung für Alzheimer in sich trugen. Den Mäusen wurden RFID - Chips implantiert, und sie lebten in einer semi-natürlichen Umgebung mit gesunden Mäusen zusammen. Ziel des Experimentes war es herauszufinden, ob an Alzheimer erkrankte Mäuse in einer semi-natürlichen Umgebung nicht so stark erkranken wie Mäuse in einem normalen Käfig. Semi-natürliche Umgebungen sind Käfige, in denen den Mäusen Spielsachen und Objekte zum Anregen der geistigen Fähigkeiten gegeben werden (beispielsweise Zeitungsschnipsel oder kleine Bälle). Insgesamt wurde mit 83 Mäusen experimentiert, von denen 22 mit Alzheimer infiziert waren. Die Mäuse wurde über acht Monate hinweg beobachtet und zu jedem Zeitpunkt lebten ungefähr 20 Mäuse im Käfig. Der Käfig wurde mit 29 RFID-Sensoren ausgestattet, welche jeweils an strategischen Plätzen, wie beispielsweise an Tränken oder an Enden von Röhren, aufgestellt wurden. Die Sensoren wurden ausgelöst, wenn sich eine Maus auf weniger als 3 cm einem Sensor näherte, und dies wurde in einer Datenbank festgehalten. Die Datenbank enthält insgesamt 8 027 645 millisekundengenaue Einträge, die jeweils die Maus und den Sensor identifizieren und den jeweiligen Zeitpunkt des Eintrags vermerken. Zusätzlich ist die Position der Sensoren im Käfig und der Gesundheitszustand jeder Maus bekannt. Durch Abgrenzung verschiedener Bereiche und Überwachung der Verbindungen zwischen den Bereichen erhält man mit nur 29 Sensoren ein ziemlich umfassendes Bild von den Bewegungen der Mäuse. In der Abbildung 1.1 ist der Käfig zu sehen, wobei man an den Kabeln die RFID-Sensoren erkennen kann. Es wurden keine Futterstellen überwacht, so dass keine Aussagen über die Fressgewohnheiten der Mäuse getroffen werden können.

Das Ziel der Datenanalyse war es, Indizien zu finden, anhand derer man auf die Krankheit bzw. Gesundheit der Mäuse schließen kann und das Verhalten der Mäuse graphisch darzustellen. Beispielsweise wurde die zurückgelegte Distanz jeder Maus untersucht und damit versucht den Gesundheitszustand unterscheiden zu können. Zusätzlich ist es interessant, Aussagen über die Trinkgewohnheiten von kranken Mäusen zu treffen, schließlich sind beide Varianten möglich: die Mäuse trinken mehr, weil sie vergessen haben, wann sie zuletzt getrunken haben, oder sie trinken weniger, weil Alzheimer das Durstgefühl negativ beeinflusst. Ferner ist die Miteinbeziehung der räumlichen Komponente sehr interessant, da sich beispielsweise das Revierverhalten verändern könnte. Durch die sporadische Platzierung von Sensoren ist keinerlei Wissen darüber vorhanden, was die Mäuse zwischen zwei Sensorbesuchen gemacht haben. Sie könnten sowohl auf direktem Weg als auch auf Umwegen zum nächsten Sensor gelangt sein. Da die Datengrundlage für alle Umwege fehlte, wurde zum Abschätzen der Strecken angenommen, dass sich die Mäuse immer auf direktem Wege zum nächsten Sensor begeben haben.

Zusätzlich kam erschwerend hinzu, dass der eigentliche Experimentaufbau auf eine andere Fragestellung hinzielte. Schließlich war die dem Experiment zu Grunde liegende Frage, ob Alzheimer - Mäuse in semi - natürlichen Umgebungen nicht so stark erkranken wie in norma-

Abbildung 1.1: Die semi-natürliche Umgebung des Experimentes war mit RFID-Sensoren ausgestattet. An den Kabeln erkennt man die Position der RFID-Sensoren, welche die Bewegungen der Mäuse aufgezeichnet haben. [13]

len Käfigen. Bei der Obduktion der kranken Mäuse aus semi-natürlichen Umgebungen wurden wesentlich weniger Ablagerungen im Gehirn gefunden als bei kranken Mäusen aus normalen Käfigen. Anhand der Stärke der Ablagerungen kann auf die Schwere des Krankheitsverlaufes geschlossen werden. Somit sind bei den vorliegenden Daten die krankheitsbedingten Verhaltensänderungen nicht so stark ausgeprägt und diese schwache Ausprägung erschwert die Unterscheidung zwischen kranken und gesunden Mäusen. Ferner waren in den letzten drei Monaten des Versuchs keine kranken Mäuse mehr im Käfig vertreten und von den 83 Mäusen waren auch nur 22 genetisch verändert. Durch diese Einschränkungen können die Analysen natürlich nicht vollständig verallgemeinerbar sein, aber sie liefern interessante und auch reproduzierbare Erkenntnisse über die Auswirkungen einer bis heute unheilbaren Krankheit: Alzheimer.

Die Forschung an Mäusen ist ein wichtiger Schritt bei der genauen Analyse von Krankheiten. Falls diese Ergebnisse auf den Menschen übertragbar sind, so könnten die Analysen der Daten interessante Aspekte aufwerfen. Bisherige Verhaltensforschungen bestanden beispielsweise darin, dass sich Menschen bemühen mussten, das Verhalten von Versuchsobjekten akribisch und perfekt zu protokollieren. Mit der heutigen Technik kann genau diese Arbeit, die keinen wirklichen menschlichen Verstand benötigt, übernommen und perfektioniert werden. Es wird jedoch fast immer nötig sein, einen Menschen mit der genauen Analyse der Daten zu beschäftigen. Der Computer kann mit Algorithmen nur bisher bekannte Muster auffinden, neue Muster sind entweder dem Menschen schon bekannt oder müssen vom Menschen daraufhin untersucht werden, ob sie für den Anwendungskontext relevant sind. Der Verstand des Menschen ist dem Computer immer noch überlegen, gerade wenn es um das Auffinden von

neuen Mustern geht. Verbindet man die Stärken des Computers – das Verarbeiten und Darstellen großer Datenmengen – mit dem „schärfsten Schwert“ des Menschen, so kann am besten aus dem Vorhandensein von Computern profitiert werden.

2 Related Work

2.1 Analyse von Tierbewegungen

Bei der Analyse von Bewegungs- und Verhaltensmustern von Labormäusen [12] wurde vor allem die Wiedergabe der Bewegungen forciert. Dazu wurde ein Computermodell des Käfigs erstellt und eine oder mehrere Mäuse konnten mittels einer Animation der Bewegung verfolgt werden. Dadurch kann sich der Verhaltensbiologe einzelne Geschehnisse mehrmals anschauen, die er in Echtzeit einmal und wahrscheinlich nur unvollständig mitbekommen hätte. Zusätzlich wurden einfache Analysemethoden implementiert, die das Verhalten der Mäuse beschreiben sollen. Beispielsweise kann für jeden Tag die Anzahl der benutzten Bereiche des Käfigs und die Anzahl der ausgelösten Sensoren bestimmt werden. Außerdem kann für jeden Käfigbereich berechnet werden, wie lange eine Maus sich dort aufhält und ob sie sich dort mit anderen Mäusen zusammen aufhält. Dieser Ansatz sollte durch den Einsatz der RFID - Technologie die Protokollierung der Mausaktivitäten durch Menschen unnötig machen und ein Wiederabspielen des Protokolls ermöglichen. Er zielte nicht so stark darauf ab, kranke und gesunde Mäuse zu unterscheiden. Es wird vorgeschlagen, eine Klassifikation durchzuführen, um kranke und gesunde Mäuse zu unterscheiden, aber wie diese genau aussehen soll, wird nicht besprochen. Die in meiner Bachelorarbeit verwendeten und analysierten Daten stammen aus demselben Experiment, mit dem sich diese Publikation befasst hat.

Eine aktuelle Forschungsarbeit zur Aufzeichnung und Analyse des täglichen Verhaltens von Kühen mittels eines lokalen Positionierungssystems [6] zeichnet die Bewegungen der Kühe mittels Triangulation per Radar auf. Die Verwendung eines radargestützten Positionierungssystems hat den Vorteil, stetige Positionen zu liefern, und bietet eine Genauigkeit von bis zu 50 Zentimetern. Jedoch wird die Auflösung im Bereich von Metallgegenständen auf der Koppel, wie beispielsweise Tränken oder Futterstellen, negativ beeinflusst. Dies erschwert eine Analyse an den für die Untersuchung des Sozialverhaltens wichtigen Stellen, schließlich kann man soziale Interaktionen besonders an Futterstellen gut beobachten. Besonderer Schwerpunkt dieser Forschungsarbeit lag in der Entwicklung der technischen Infrastruktur, die ein einfach auf- und abzubauendes und effektives Aufzeichnen von Bewegungen ermöglicht. Schließlich wurde eine automatische Analyse der Daten vorgenommen, bei der Zeitpunkte sozialer Interaktionen bestimmt werden sollten. Als soziale Interaktion wurde hierbei das Verscheuchen einer Kuh A durch eine ranghöhere Kuh B angesehen. Dieses Bewegungsmuster zeichnet sich durch die Eigenschaft aus, dass Kuh A eine gewisse Zeit still an einer Position steht und Kuh B sich solange ihr nähert, bis Kuh A die Position verlässt und Kuh B nun an dieser Stelle stehen bleibt. Um die automatische Analyse zu validieren, wurden diese sozialen Interaktionen von Menschen beobachtet und der Zeitpunkt festgehalten. Von den 25 beobachteten Interaktionen wurde von der automatischen Analyse keine einzige entdeckt. Dies wird durch die kurze

Dauer und den Ort (meistens bei den schlecht überwachten Bereichen wie beispielsweise den Tränken) der Interaktionen begründet.

2.2 Visualisierung von Zeitreihen

Die TimeWheel - Technik [16] bietet eine Repräsentation multivariater Daten mittels mehrerer Achsen. Dabei wird die Zeitachse bildschirmmittig dargestellt und die anderen, zeitabhängigen Dimensionen werden nun zirkular um die Zeitachse herum angeordnet. Anschließend werden für jeden Zeitpunkt, für den Daten vorhanden sind, Linien von dem Zeitpunkt auf der Zeitachse zu den jeweiligen Punkten auf den Attributsachsen gezeichnet. Durch Interaktion kann das TimeWheel um die eigene Achse gedreht werden, und es besteht die Möglichkeit, Bereiche von Achsen zu vergrößern und somit auch nur bestimmte Zeitabschnitte zu untersuchen. Dieser Ansatz bietet sich durch das Zeichnen von Linien eher für Datensätze an, welche sehr stark mit der Zeit korrelieren. Bei dieser Technik gibt es einen hohen Grad der Überlappung, da sich die Linien zu den einzelnen Dimensionen leicht überdecken können. Ferner sind periodische Aktivitäten mit dieser Technik nur schwer auffindbar, weil die Zeitdimension nur linear dargestellt wird.

Bei der Analyse von Finanzdaten ist die Verwendung von zweidimensionalen Colormaps ein Ansatz, der in [17] verfolgt wurde. Beim zweidimensionalen Colormap erhält man für zwei Indexwerte eine Farbe. In der beschriebenen Technik gibt die erste Dimension den absoluten Gewinn oder Verlust an, den ein Käufer einer Aktie gemacht hätte. Mittels der zweiten Dimension kann die relative Entwicklung einer Aktie im Bezug zum gesamten Markt untersucht werden. Es wird also gezeigt, ob sich eine Aktie im Vergleich zum Markt besser oder schlechter verhalten hat. Mittels dieses Colormaps – in Abbildung 2.1 gezeigt – kann beispielsweise eine gewinnbringende und dem Markt gegenläufige Aktie entdeckt werden. Visualisierungen, welche diesen Colormap verwenden, sind für den Laien jedoch nicht intuitiv verständlich, und es erfordert etwas Übung, um den Farben sofort Bedeutung zuzuordnen. Für die Visualisierung der Mausbewegungen wird im späteren Verlauf der vorliegenden Arbeit ein dreidimensionaler Colormap verwendet, um die zeitliche Dimension darzustellen.

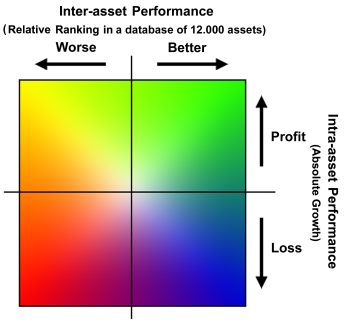


Abbildung 2.1: Der zweidimensionale Colormap, welcher zur Analyse von Aktien verwendet wurde. [17]

Da die Mausdaten eine sehr große und lange Zeitreihe sind, muss eine geeignete Visuali-

sierung für lange Zeitreihen verwendet werden. Die Recursive Patterns [10] sind eine platzausfüllende Visualisierung für große Datenmengen und bieten sich daher an. In Abbildung 2.2 wird gezeigt, auf welche Weisen der Benutzer die Datenpunkte anordnen kann. Je nach Parameterwahl können unterschiedliche Anordnungen, wie beispielsweise eine Zeilen- oder Spaltenanordnung erzielt werden. Die Recursive Patterns ermöglichen daher eine der Semantik in den Daten entsprechende Anordnung der Datenpunkte. Beispielsweise kann man eine stundengenaue Zeitreihe in Tage, Monate und Jahre auch visuell unterteilen. Die Anwendung dieser Technik ermöglicht somit das schnelle Auffinden von zeitabhängigen Mustern, wie beispielsweise Wochenenden oder Tag- und Nachtzyklen. Es ist also möglich, anwendungsspezifisches Wissen in die Visualisierung einfließen zu lassen. Daher wird unter anderem auch dieser Ansatz für die Analyse der Mausbewegungen verwendet.

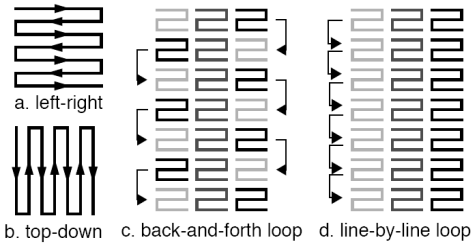


Abbildung 2.2: Hier werden die möglichen Anordnungen von Datenpunkten mittels der Recursive Patterns gezeigt. Durch eine geeignete Parameterwahl kann die Semantik der Daten berücksichtigt werden. [10]

Für große Zeitreihen, die eine Periodizität aufweisen, bietet sich auch die SpiralGraph - Visualisierung [4] an. Bei diesem Verfahren werden die einzelnen Datenpunkte spiralförmig angeordnet, wobei für die äußeren Datenpunkte mehr Platz zur Verfügung steht als für Datenpunkte im Inneren. Mittels dieser Anordnung kann pro Umdrehung beispielsweise eine Woche visualisiert werden. Um das periodische Muster erkennen zu können, muss die Krümmung der Spirale jedoch richtig gewählt werden. Anders als beim Recursive Pattern kann keine verschachtelte Anordnung der Zeitreihe erreicht werden, dafür kann der Benutzer die Krümmung der Spirale interaktiv (oder auch automatisch) einstellen. Somit bietet sich die Visualisierung vor allem für Zeitreihen mit nur einer einzigen Periode an, falls aber in den Daten zusätzliche Unterperioden sind, ist die rekursive Anordnung der Recursive Pattern von Vorteil.

2.3 Visualisierung von zeitabhängigen, räumlichen Daten

Die gleichzeitige Darstellung zeitlicher und räumlicher Dimensionen ist die größte Herausforderung bei der Visualisierung der Daten. Bisherige Ansätze, die sich genau mit diesem Problem beschäftigt haben, sind beispielsweise der GeoTime - Ansatz [9] und die Space-Time Cube Technik [11]. Bei beiden wird die Zeit den zweidimensionalen räumlichen Daten als dritte Dimension hinzugefügt. Die zu Grunde liegende Karte wird in der gerade aktuellen Zeitebene dargestellt und die Bewegungen der Objekte werden als Linien im dreidimensionalen Raum eingezeichnet. Durch diesen Ansatz hat der Benutzer den Überblick über benach-

barte Zeitpunkte und kann Bewegungen über die Zeit gut verfolgen. Auf diese Weise kann beispielsweise ein Tagesverlauf einer Person dargestellt werden. Dieser Ansatz skaliert jedoch nicht mit der Anzahl der Objekte, die mittels Linien dargestellt werden sollen. Weil zu jeder Zeit ungefähr 20 Mäuse aktiv sind und insgesamt für manche Mäuse bis zu 8 Monate Zeitraum visualisiert werden müssten, kann man diese Art der Visualisierung für die Analyse der Mausbewegungen eher nicht verwenden. Zusätzlich kann eine zweidimensionale Ausgabe von dreidimensionalen Modellen die Analyse erschweren, da Linien überlappen und sich gegenseitig verdecken können. Ferner bietet dieser Ansatz keinen Überblick über häufige Bewegungsmuster und eignet sich somit kaum für die Unterscheidung von Mäusen auf Grund ihrer verschiedenen Bewegungsmuster.

Auch bei der Visualisierung von Daten aus einer einjährigen Überwachung eines großen Bürobereichs [8] mussten räumliche und zeitliche Dimensionen gleichzeitig berücksichtigt werden. Da die Überwachung vor allem mittels Bewegungssensoren und nur wenigen Videokameras durchgeführt wurde, konnte die Zuordnung einer Bewegung nicht immer präzise erfolgen. Dieses Problem tritt durch den Einsatz der RFID - Technologie im Laborversuch mit den Mäusen nicht auf, da jeder RFID - Chip eine eindeutige Identifikationsnummer trägt. Zusätzlich zielte der Ansatz auf das effiziente Auffinden wichtiger Zeitpunkte, beispielsweise das Auftreten eines Bewegungsmusters. Außerdem sollte sich der Benutzer entsprechende Videoszenen anschauen und die Bewegungen als Animation ablaufen lassen können. Für die Aufgabenstellung sind diese Funktionalitäten nicht notwendig, da es nicht um einzelne Szenen oder um schon bekannte Bewegungsmuster geht, sondern um die Klassifikation des Gesundheitszustandes von Mäusen. Schließlich war die Aufgabe herausfinden, was kranke Mäuse auszeichnet und was sie für ein Bewegungsprofil haben.

Zeitreihen mit räumlichem Bezug können auch mit den Lexis Pencils [3] dargestellt werden. Hierbei werden bleistiftähnliche geometrische Objekte verwendet, auf deren Seitenflächen die zeitabhängigen Variablen gezeichnet werden. Hierbei verläuft die Zeitachse von der Spitze zum Ende des Bleistiftes. Die Bleistifte können anschließend in einem dreidimensionalen Raum entsprechend der räumlichen Position angeordnet werden und bieten somit die Möglichkeit, heterogene Daten zu visualisieren. Ein Problem bei dieser Technik ist, dass eigentlich nur die dem Benutzer zugewandte Seite des Bleistiftes verwendet werden kann, da die hintere Seite nicht sichtbar ist. Zusätzlich verhindert die lineare Darstellung der Zeitreihen Periodizitäten einfach zu erkennen.

3 Zeitliche Analyse der Sensordaten

Dieser Teil der Arbeit beschäftigt sich mit der zeitlichen Analyse der Mausdaten ohne Berücksichtigung der räumlichen Dimension. Die dabei zugrunde liegende Fragestellung ist, ob man auch schon allein ohne die räumliche Dimension Aussagen über die Daten treffen und diese Aussagen auch statistisch validieren kann. Das Hauptproblem bei diesen Daten besteht in der Verknüpfung von räumlicher und zeitlicher Dimension. Wenn also nur die Zeitdimension allein schon für die Datenanalyse und Klassifikation der Mäuse nach Gesundheit ausreicht, kann ein großer Aufwand gespart und dem Benutzer zusätzlich eine einfache und leicht verständliche Visualisierung präsentiert werden.

Damit die alleinige Berücksichtigung der zeitlichen Dimension überhaupt Sinn macht, wurden neue Werte aus den Daten generiert. So wurde beispielsweise die gelaufene Strecke pro Stunde berechnet und die Häufigkeit eines Tränkenbesuchs registriert. Mittels Visualisierung und Auswertung dieser Daten können so auch schon interessante Einblicke in das Mausverhalten gewährt werden. Um zusätzliche Aussagen über das generelle Verhalten von kranken und gesunden Mäusen treffen zu können, wurden aus den Mausdaten acht Prototyp-Mäuse generiert, die jeweils eine der Kombinationen aus Geschlecht und Gesundheitszustand präsentieren. Diese Prototypen stehen also für gesunde bzw. kranke und männliche bzw. weibliche Mäuse und jeweils noch die vier Kombinationen zwischen den zwei Attributen. Diese Prototypen wurden durch eine Durchschnittsbildung über die entsprechenden Mausaktivitäten berechnet.

Bei der Durchschnittsberechnung musste aber die unterschiedliche Bevölkerungsstruktur des Käfigs berücksichtigt werden: zu jedem Zeitpunkt konnte die Zusammensetzung der Käfigbewohner variieren. Beispielsweise sind am Anfang des Versuchs fünf weibliche Mäuse im Käfig und sieben Monate später beim Ende der Datenaufzeichnung keine einzige mehr. Somit muss ein angepasster Durchschnitt für die Berechnung der Prototyp-Aktivitäten ausgedacht und anschließend verwendet werden.

$$\bar{x}(\text{Zeitpunkt } t, \text{Eigenschaft } e) = \sum_{\text{Mäuse } m \text{ mit Eigenschaft } e} \frac{\text{Wert}(m,t)}{\text{Anzahl Mäuse}(t,e)} \tag{3.1}$$

Wie in Formel 3.1 gezeigt, wird der Durchschnitt der Werte nicht mittels eines festen Quotienten berechnet, sondern durch Berücksichtigung der zum Zeitpunkt t vorhandenen Mäuse, welche die gewünschte Eigenschaft e besitzen. Da dieser Ausdruck nicht definiert ist, falls keine Mäuse vorhanden sind, wird in diesem Fall angenommen, dass dieser Ausdruck das Ergebnis Null hat. Durch diese Berechnung kann man also die Prototypen auch miteinander vergleichen und ist sicher, durch das Verwenden von Prototypen keine künstlich generierten Aussagen zu erzeugen.

3.1 Anwendung der Recursive Patterns

Es gibt sehr einfache Ansätze zur Visualisierung von Zeitreihen, wie beispielsweise die Darstellung mittels Liniendiagrammen. Liniendiagramme haben den Nachteil, nicht den gesamten Platz auszunutzen. Hierdurch können die Daten nicht so detailliert dargestellt werden, wie auf Grund des zur Verfügung stehenden Platzes möglich wäre. In Abbildung 3.1 sind die gelaufenen Strecken einer männlichen kranken Maus über einen Zeitraum von 101 Tagen (ungefähr 14 Wochen) dargestellt. Die gelaufenen Strecken wurden jeweils für eine Stunde aggregiert und berechnet, somit stehen $101 \cdot 24 = 2424$ Datenpunkte für die Visualisierung zur Verfügung. Durch die hohe Datendichte und gleichzeitig schlechte Ausnutzung der zur Verfügung stehender Fläche kann keine wirkliche Aussage über das Verhaltensmuster getroffen werden. Der Vergleich der Breite der Visualisierung und der Anzahl der Datenpunkte führt zu dem Ergebnis, dass pro Pixel zwei Datenpunkte dargestellt werden müssen. Dies erschwert natürlich auch das Erkennen eines Tageszeitenrythmus. Allein Sensorausfälle und globale Trends, wie beispielsweise das Abflachen der Peaks, sind deutlich sichtbar und fallen dem Betrachter sofort ins Auge.

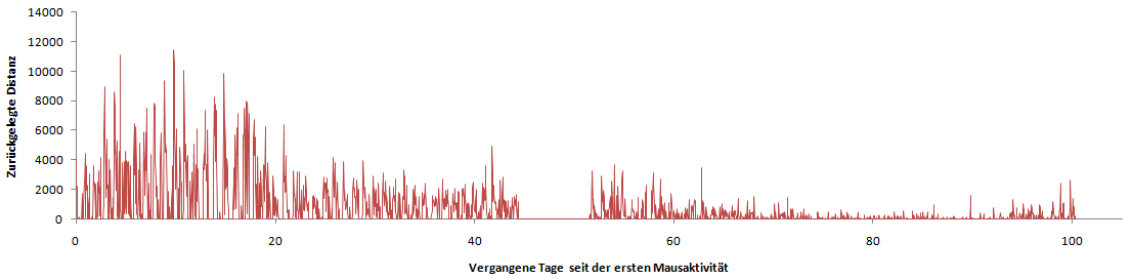


Abbildung 3.1: Hier sind die pro Stunde aggregierten Laufstrecken einer männlichen kranken Maus in einem Liniendiagramm dargestellt.

Die gelaufene Strecke wird in Längeneinheiten von 1.75 cm gemessen, wobei diese Längeneinheit durch das verwendete Koordinatensystem begründet ist. Der Käfig hat eine Länge und Breite von 1.75 Metern und wurde mit einem 100×100 Gitter in den Computer abgebildet. Die oben dargestellte Maus ist beispielsweise in einer Stunde maximal 200.12 Meter gelaufen.

Zum Vergleich der beiden Techniken wurden in Abbildung 3.2 dieselben Daten mit einem Recursive Pattern dargestellt. Jedes kleine farbige Rechteck steht für eine Stunde, wobei die Farbsättigung kodiert, wieviel die Maus gelaufen ist. Die Farbsättigung wird mittels einer linearen 0-1-Normalisierung berechnet, wobei gesättigtere Einfärbungen für eine größere Distanz stehen. Jede Zeile (bestehend aus 24 Rechtecken) steht für einen Tag von 0 - 23 Uhr und jedes große Rechteck mit 31 Zeilen steht für einen Monat.

Allein wenn man die Anzahl der darstellbaren Datenpunkte vergleicht, erkennt man den Vorteil des Recursive Pattern. Angenommen die Zeichenfläche hätte die Höhe h und die Breite b, so können bei Liniendiagrammen maximal b Datenpunkte eingezeichnet werden. Bei Recursive Pattern hingegen können $b \cdot h$ Datenpunkte visualisiert werden und zusätzlich kann der Aufbau der Recursive Patterns mittels Parametern bestimmt werden. Die hierdurch mög-

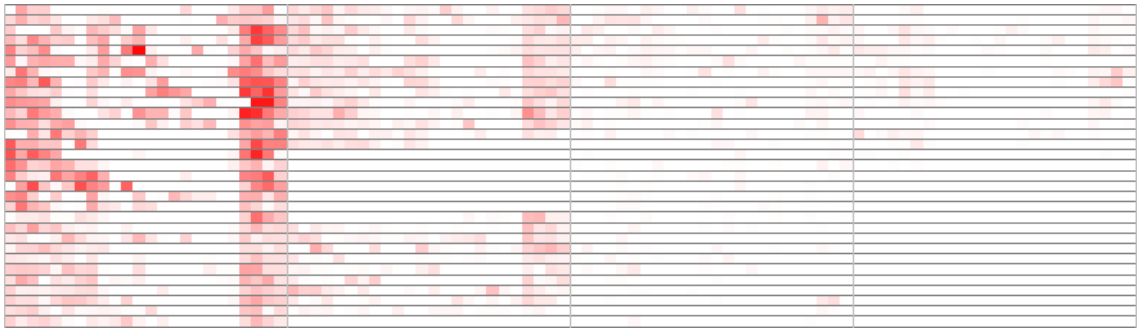


Abbildung 3.2: Die Laufstrecken einer männlichen kranken Maus mittels Recursive Pattern dargestellt. Beispielsweise ist der Tagesrythmus an der Bande, die um ca. 20 Uhr beginnt, deutlich erkennbar.

liche Gliederung der Daten in Stunden, Tage und Monate ermöglicht eine genauere Analyse der Daten und das Auffinden von vielleicht vorhandenen Unterschieden in den Bewegungsmustern oder Tagesabläufen der kranken und gesunden Mäuse. Beispielsweise ist in Abbildung 3.2 ersichtlich, dass die hier dargestellte Maus jeweils zu den Dämmerungsstunden ihre Hauptaktivitätszeiten hat, aber manchmal auch tagsüber aktiv ist. Der Vorteil gegenüber den Liniendiagrammen liegt also zum einen in dem höheren Detailgrad der Darstellung und zum anderen im Einbeziehen der zeitlichen Struktur.

3.1.1 Untersuchung der Laufstrecken

In diesem Abschnitt wird die Analyse, welche sich mit den Auswirkungen von Alzheimer auf die gelaufenen Strecken befasst, vorgestellt. Dabei werden die Laufstrecken jeder Maus pro Stunde berechnet und dies dient als Grundlage für die Analyse. Durch die oben erwähnten Maus - Prototypen können auch Aussagen bezüglich einer Mausgruppe (beispielsweise alle kranken Mäuse) getroffen werden. Dies ermöglicht erst eine genaue Abgrenzung der beiden Klassen voneinander und kann auch dazu benutzt werden, unterschiedliche Auswirkungen von Alzheimer auf männliche und weibliche Mäuse festzustellen.

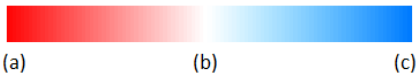


Abbildung 3.3: Verwendeter Colormap, um die unterschiedlichen zurückgelegten Strecken zu visualisieren. Bei (a) ist die zweite Maus mehr gelaufen als die erste (Differenz ist negativ), bei (b) sind beide Mäuse gleich weit gelaufen und bei (c) ist die erste Maus weiter gelaufen als die zweite (Differenz ist positiv).

Um Vergleiche zwischen den zwei Mausklassen anstellen zu können, müssen Aussagen über die Unterschiede einfach und intuitiv getroffen werden können. Dies ist aber nicht der Fall, wenn nur zwei Recursive Patterns untereinander gezeichnet werden, wobei jedes Recursive Pattern den Bewegungsablauf einer Maus visualisiert. Bei dieser Darstellung der Daten

müsste der Betrachter die zwei Farbsättigungen des gleichen Zeitpunktes im Kopf vergleichen und sich das Ergebnis merken. Folglich wäre eine solche Visualisierung noch nicht ausreichend, um eine effiziente Analyse der Daten vorzunehmen. Da vor allem interessiert, welche Maus die größere Strecke zurückgelegt hat, und nicht so wichtig ist, wie viel die Maus mehr gelaufen ist, liegt das Hauptaugenmerk auf dem Visualisieren von kategorischen Daten. Wie von Mackinlay in [14] angegeben, sollte man, von der Position abgesehen, vor allem den Farbton für das Darstellen von kategorischen Daten verwenden. Ferner kann noch dargestellt werden, um wie viel eine Maus weiter gelaufen ist, wenn zusätzlich die Farbsättigung für die Visualisierung verwendet wird. Somit wird die Differenz der zurückgelegten Strecken pro Zeitpunkt berechnet und auf ein zweifarbiges Colormap übertragen. Um einen intuitiven Zugang zu dem zweifarbigen Colormap zu bieten, nimmt die Färbung eines Rechtecks jeweils den Farbton der Maus an, welche die weitere Strecke zurückgelegt hat. Dieser Colormap, in Abbildung 3.3 dargestellt, bietet zum einen den schnellen Überblick, welche Maus wann aktiver war, und zum anderen auch einen detaillierten Vergleich der jeweiligen zurückgelegten Strecken.

Das in Abbildung 3.4 gezeigte Tool ermöglicht den Vergleich von zwei Mäusen miteinander. Im abgebildeten Beispiel wurden die Prototypen der weiblichen gesunden und weiblich kranken Mäuse dargestellt. Durch die gleichzeitige Anzeige zweier Mäuse können erste Aussagen über das Verhalten von kranken und gesunden Mäusen getroffen werden. Die Visualisierung besteht aus drei Zeilen: in den ersten beiden Zeilen ist das jeweilige Recursive Pattern für die zu vergleichenden Mäuse dargestellt und in der letzten Zeile wird das oben vorgestellte Differenzbild der beiden Recursive Patterns gezeichnet. Da die Mäuse im Allgemeinen nicht gleich lang gelebt haben, sind die Zeiträume, für die Daten vorhanden sind, nicht gleich groß. Ein Vergleich macht aber nur dann Sinn, wenn für jede Maus ein Wert vorhanden ist. Somit berechnet das Programm den Schnitt der beiden Zeiträume und visualisiert nur die Zeitpunkte, die innerhalb des Schnittes liegen. Um die Farbsättigung zu berechnen, muss eine Normalisierung der Daten erfolgen, wobei das Maximum der Zeitreihen a priori nicht bekannt ist. Daher sucht das Programm den maximalen Wert aus der Vereinigung der Zeitreihen beider Mäuse und verwendet diesen Wert für die anschließende Normalisierung. Die verwendeten Colormaps wurden mit dem Colormap-Tool [2] erzeugt, welches die zu verwendenden Farben als RGB - Farbwerte ausgibt. Dies ermöglicht eine einfache Integration der generierten Colormaps in das Programm.

3.1.2 Untersuchung der Trinkhäufigkeit im Bezug zur Laufstrecke

Bei dieser Analyse sollte festgestellt werden, ob sich der Gesundheitszustand auf die Trinkhäufigkeit auswirkt. Die Annahme war, dass verstärkte Aktivität, beispielsweise längere Laufdistanzen, ähnlich wie beim Menschen zu einem erhöhten Flüssigkeitsbedarf führt. Mittels Verwendung der Maus-Prototypen sollte ein möglicher Unterschied zwischen dem Trinkverhalten von gesunden und kranken Mäusen aufgezeigt werden. Ferner sollte zur Validierung der Ergebnisse ein Unterschied zwischen männlichen und weiblichen Mäusen feststellbar sein, da der Flüssigkeitshaushalt bei den verschiedenen Geschlechtern unterschiedlich ist.

Für die Analyse wurde ein Wert generiert, welcher die Trinkhäufigkeit im Bezug zur zurückgelegten Strecke in Relation bringt. Dieser Wert sollte, um ein intuitives Verständnis zu

Abbildung 3.4: Vergleich von gesunden (blau) und kranken (rot) Mäusen, wobei die Sättigung, welche die zurückgelegte Strecke repräsentiert, mittels einer Quadratwurzel-Normierung berechnet wurde.

ermöglichen, groß sein, falls häufig getrunken und eine kurze Strecke zurückgelegt wurde, und klein sein, falls selten getrunken und eine lange Strecke zurückgelegt wurde.

$$\text{tränkenProStrecke}(\text{Maus } m, \text{Zeitintervall } t) = \frac{anzahlWasserstellen(m,t)}{gelaufeneStrecke(m,t)} \tag{3.2}$$

Der Ausdruck in Gleichung 3.2 erfüllt die geforderten Eigenschaften und ist einfach zu berechnen. Dieser Wert kann wie die zurückgelegten Strecken mit dem oben vorgestellten Tool visualisiert werden. In Abbildung 3.5 ist der Vergleich zwischen kranken und gesunden Mäusen zu sehen. Das Differenzbild ist sehr aussagekräftig: man erkennt, dass mittels dieser Herangehensweise an den generierten Datenwert keine Aussage über die Unterschiede zwischen kranken und gesunden Mäusen getroffen werden kann.

Da die visuelle Untersuchung dieses Wertes keine Aussage ergab, wurde eine statistische Analyse durchgeführt. In der folgenden Tabelle sind die jeweiligen Mittelwerte für jede Mausgruppe eingetragen und können so miteinander verglichen werden. In der letzten Spalte sind die Werte für jedes Geschlecht eingetragen und in der letzten Zeile die Werte für jeden Gesundheitsstatus; der Eintrag ganz rechts unten gibt den Mittelwert über alle Mäuse an.

tränkenProStrecke	gesund	krank	
männlich	0.00692359	0.00626108	0.00675281
weiblich	0.00571020	0.00497311	0.00556161
	0.00633801	0.00566885	0.00614030

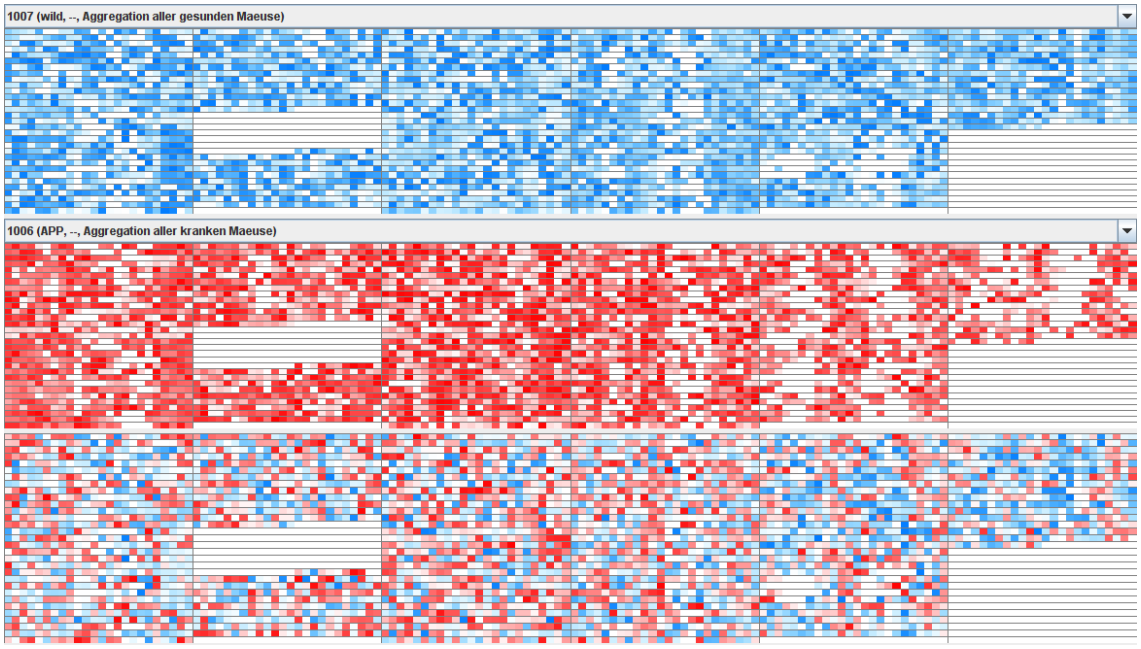


Abbildung 3.5: Visualisierung der Trinkhäufigkeit im Bezug zur gelaufenen Strecke bei gesunden (blau) und kranken (rot) Mäusen. Trotz einer linearen Normalisierung ist kaum ein Unterschied erkennbar.

Anhand dieser Werte erkennt man, dass männliche Mäuse im Bezug zur gelaufenen Strecke mehr trinken als weibliche Mäuse. Allerdings ist viel wichtiger, dass gesunde Mäuse im Bezug zur zurückgelegten Strecke mehr trinken als kranke Mäuse. Ferner ist die Auswirkung der Krankheit auf die Trinkhäufigkeit bei weiblichen Mäusen stärker ausgeprägt als bei den männlichen Mäusen. Betrachtet man den Durchschnitt über alle Mäuse (0.00556161), so kann man aus diesem Wert ausrechnen, dass eine Maus durchschnittlich alle 2.85 Meter eine Tränke aufsucht.

Nachdem nun erst die zweite Analyse der Daten eine interessante Aussage erbracht hat, ist wichtig zu überprüfen, ob die Aussage, dass gesunde Mäuse mehr trinken als kranke, überhaupt statistisch relevant ist. Zum Vergleich zweier Messreihen gibt es eine standardisierte Zufallsvariable, aus welcher man die Sicherheit der statistischen Relevanz berechnen kann (vgl. Kapitel 1.1.3 „Tests und Konfidenzintervalle bei unbekannten und ungleichen Varianzen [...] “ in [7]). t kann mit der Formel aus 3.3 berechnet werden, wobei $\bar{x}$ für den Mittelwert, n für die Anzahl der Elemente und S für die Stichprobenvarianz steht. Anschließend wird das maximale α bestimmt, bei dem das α-Quantil der Normalverteilung noch kleiner als t ist. Das Verwenden der Normalverteilung anstatt der t-Verteilung ist durch die große Zahl von Datenelementen erlaubt, da die t-Verteilung für eine Elementanzahl größer als 120 die Normalverteilung approximiert.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \text{ mit } S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2 \text{ und } k \in \{1,2\} \tag{3.3}$$

Dieser statistischer Test wurde für die Aussage über den Einfluss des Geschlechts auf die Trinkhäufigkeit und für die Aussage, dass gesunde mehr trinken als kranke Mäuse, durchgeführt. Für die erste Aussage erhält man einen Wert von $t = 24.13768$ und für die zweite Aussage einen Wert von $t = 13.75452$. Da man das maximale α nicht so einfach bestimmen kann, wurde als Referenzwert das α-Quantil der Normalverteilung für $\alpha = 0.999999999999$ ausgerechnet. Für dieses Quantil erhält man einen Wert von $Q_\alpha = 7.03449$, somit ist die statistische Relevanz der Daten noch höher als das hier angegebene α. Man kann also mit recht hoher Wahrscheinlichkeit davon ausgehen, dass beide Aussagen statistisch signifikant sind.

3.2 Hierarchisches Clustering

Neben den oben vorgestellten Analysen, war auch interessant zu wissen, ob man allein auf Grund der gelaufenen Distanzen zwischen kranken und gesunden Mäusen unterscheiden kann. Dazu wurde für jede Maus ein durchschnittlicher Tagesablauf bezüglich der gelaufenen Strecken berechnet. Es wurde also für jede Stunde des Tages ein durchschnittlicher Aktivitätswert berechnet, zum Beispiel der Durchschnitt über alle Distanzen, die zwischen 0 und 1 Uhr zurückgelegt wurden. Somit konnte ein Clustering auf einem 24 - dimensionalen Vektor durchgeführt werden. Die durchschnittlichen Werte wurden anschließend über alle Mäuse hinweg mit der Quadratwurzelnormierung auf das Intervall [0, 1] normalisiert. Für das hierarchische Clustering wurde eine Manhattan - Distanz verwendet, da beispielsweise ein euklidischer Abstand bei zurückgelegten Distanzen nicht sinnvoll wäre. Das hier vorgestellte hierarchische Clustering wurde in R [5] mittels der vorhandenen Funktion *hclust* durchgeführt unter der Verwendung des average linkage Clusterings und anschließend wurde dort auch das Dendrogramm erstellt.

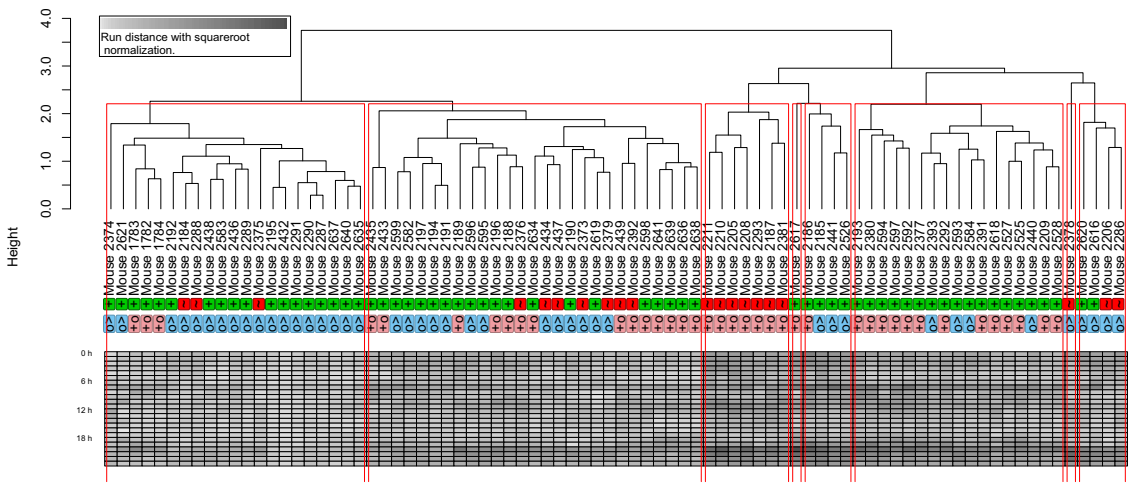


Abbildung 3.6: Hier ist das Cluster - Dendrogramm des hierarchischen Clusterings dargestellt. Die verwendeten Feature-Vektoren – bzw. der durchschnittliche Tagesablauf – wurden unter die jeweilige Maus gezeichnet. Die rote und grüne Einfärbung zeigen die Krankheit bzw. Gesundheit der Maus an und eine rosa bzw. blaue Einfärbung kodiert das Geschlecht.

In Abbildung 3.6 ist das Ergebnis des hierarchischen Clusterings in Form eines Dendrogramms dargestellt. Mittels der Einfärbung rot / grün wurde eingezeichnet, ob die Maus krank oder gesund war, und mit rosa / blau wurde das Geschlecht gekennzeichnet. Unter dem Dendrogramm sind die verwendeten Feature-Vektoren dargestellt, wobei der Tag oben bei 0 Uhr anfängt und unten bei 23 Uhr aufhört. Mittels der roten Linien wurden insgesamt 8 Cluster gebildet, um die Güte des Clusterings ersichtlich zu machen.

Wenn man sich das Clustering - Ergebnis anschaut, so fallen vor allem – von links gezählt – das dritte Cluster (nur weiblich kranke Mäuse) und das sechste Cluster (nur gesunde Mäuse)

auf. Da die anderen Cluster nicht so gut sind, stellte sich die Frage, wie wahrscheinlich es ist, dass diese Cluster zufällig zusammengewürfelt wurden. Um dieser Frage nachzugehen, wurde die vereinfachende Annahme getroffen, dass die Zuordnung zu einem Cluster einem Zufallsexperiment mit Ziehen ohne Zurücklegen und ohne Beachtung der Reihenfolge gleicht. Um die Wahrscheinlichkeit einer zufälligen Ziehung zu berechnen, kann nun die hypergeometrische Verteilung (siehe Formel 3.4) genutzt werden, da sie dieses Experiment genau modelliert. Bei dieser Formel ist N die Anzahl aller Objekte, n die Größe der Stichprobe, M die Anzahl aller Objekte mit gewünschter Eigenschaft und x ist die Anzahl der Objekte in der Stichprobe mit der gewünschten Eigenschaft.

$$p = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \tag{3.4}$$

Exemplarisch wäre für das dritte Cluster: $N = 83, n = 7, M = 10, x = 7$. Mit diesen Werten kann man nun die Wahrscheinlichkeit von $p = 2.649 \cdot 10^{-8}$ ausrechnen, dass das dritte Cluster (nur kranke weibliche Mäuse) zufällig gezogen wird. Für das sechste Cluster mit ausschließlich gesunden Mäusen erhält man eine Wahrscheinlichkeit von $p = 0.0029$. Somit kann man mit recht großer Wahrscheinlichkeit davon ausgehen, dass dieses Clustering nur auf Grund der Daten und nicht zufällig resultiert.

3.3 Ergebnisse

In diesem Abschnitt werden die Ergebnisse der einzelnen oben beschriebenen Analysetechniken vorgestellt. Durch den Vergleich von Mausprototypen ist es möglich, auch generelle Aussagen über die Auswirkungen der Alzheimerkrankheit auf Mäuse zu treffen.

Beim ersten Vergleich der Auswirkungen auf die verschiedenen Geschlechter mit dem in Abschnitt 3.1.1 vorgestellten Tool ergab sich ein deutlicher Unterschied zwischen weiblichen und männlichen Mäusen. Laut Differenzbild laufen weibliche Mäuse mehr als ihre männlichen Artgenossen, wobei dies ein Indikator für die Validität der Datenanalyse ist. Betrachtet man das Territorialverhalten der Mäuse, so stellt man fest, dass Männchen ihr Territorium gegen andere männliche Mäuse verteidigen und damit nicht so häufig an weit auseinander liegenden Sensoren vorbeikommen. Hingegen werden weibliche Mäuse von allen Männchen geduldet und können sich somit freier im Käfig bewegen. Während diese Aussage für die Untersuchung von Alzheimer noch nicht interessant ist, konnte auch eine wichtigere Erkenntnis gewonnen werden. Männliche und weibliche Mäuse reagieren nämlich ähnlich auf die Infektion mit Alzheimer. Dies wurde vor allem deutlich, als die Auswirkungen für jedes Geschlecht einzeln untersucht wurden. Dabei wurden zuerst die weiblich kranken mit den weiblich gesunden und anschließend die männlich kranken mit den männlich gesunden Mäusen verglichen. Die fast gleichen Befunde bei den zwei Vergleichen deuteten auf ein nicht sehr unterschiedliches Krankheitsbild hin.

Beim Vergleich der weiblichen Mäuse gab es eine zeitabhängige Veränderung der zurückgelegten Stecken. Während der ersten drei Monate rannten die kranken weiblichen Mäuse mehr als die gesunden und nach den drei Monaten waren die gesunden aktiver als die kranken Mäuse. Ferner waren die kranken Mäuse fast nur außerhalb der Dämmerung aktiver als

die gesunden Mäuse. Daran erkennt man, dass die kranken Mäuse vom gleichmäßigen Tagesrhythmus abwichen. Die kranken Mäuse waren somit fast den ganzen Tag aktiv und die regelmäßigen Aktivitätsbanden (beispielsweise in Abbildung 3.2 gezeigt) nicht so deutlich sichtbar.

Kranke männliche Mäuse wichen von ihrem Tagesrhythmus ähnlich stark ab wie die kranken weiblichen Mäusen. Der nachlassende Aktivitätsdrang erfolgte innerhalb von zweieinhalb Monaten. Der einzige Unterschied zwischen den beiden Geschlechtern liegt darin, dass kranke weibliche Mäuse vor allem in den Ruhepausen der gesunden Mäuse aktiver waren, wohingegen die männlichen kranken Mäuse über den gesamten Tag hinweg aktiver waren als die männlichen gesunden Mäuse. Insgesamt kann man sagen, dass weibliche und männliche Mäuse im Bezug auf die zurückgelegten Laufdistanzen ähnlich auf Alzheimer reagieren.

Bei der Analyse der Trinkhäufigkeit im Bezug zur gelaufenen Strecke ergab sich mit sehr hoher Wahrscheinlichkeit eine signifikante Unterscheidung zwischen kranken und gesunden Mäusen. Durch die durchgeführte statistische Validierung kann man davon ausgehen, dass kranke Mäuse im Bezug zur gelaufenen Strecke weniger trinken als gesunde Mäuse. Bei zusätzlicher Betrachtung der in Abschnitt 3.1.2 vorgestellten Zahlen fällt die unterschiedliche Auswirkung auf weibliche und männliche Mäuse auf. Die Verringerung der Trinkmengen ist bei weiblichen Mäusen stärker ausgeprägt als bei männlichen Mäusen. Berechnet man die Signifikanz der unterschiedlichen Trinkhäufigkeiten von kranken und gesunden weiblichen Mäusen, so erhält man einen Wert von ca. 85 % ($t = 1.10975$). Zwischen kranken und männlichen Mäusen erhält man einen Wert von 99,99 % ($t = 4.94516$). Der etwas schlechtere Wert von 85 % ergibt sich durch eine um den Faktor 10 größere Varianz (im Vergleich zur Varianz der männlichen kranken Mäuse) bei den kranken weiblichen Mäusen, was zu einem kleineren Wert von t führt. Durch die Analyse der Trinkhäufigkeit konnte nachgewiesen werden, dass sich Alzheimer auf die Trinkgewohnheiten auswirkt, wobei eine besonders starke Veränderung bei den weiblichen Mäusen festgestellt werden konnte.

Das durchgeführte hierarchische Clustering anhand der Tagesabläufe ergab eine Zweiteilung der Daten. Ungefähr zwei Fünftel der Daten lassen sich gut anhand der durchschnittlichen zurückgelegten Distanzen unterscheiden. Durch den Vergleich mit einem Zufallsexperiment konnte gezeigt werden, dass beispielsweise das Cluster bestehend aus sieben kranken weiblichen Mäusen kaum zufällig gefunden wurde. Somit kann durch alleinige Betrachtung der zurückgelegten Strecken schon eine Unterscheidung bei vierzig Prozent der Mäuse ermöglicht werden. Dies bedeutet, dass sich die Tagesabläufe durch die Krankheit bei einigen Mäusen signifikant verändern, wobei die durchschnittlich zurückgelegten Strecken nicht als alleiniges Diskriminierungsmerkmal ausreichen. Aus diesem Grund werden im nächsten Kapitel die räumlichen Informationen in die Analyse einbezogen.

4 Räumliche Analyse der Daten

Da es nicht möglich war, die Mäuse allein mittels Aktivitätszeitreihen zu klassifizieren, werden in diesem Kapitel zwei weitere Analysemethoden vorgestellt. Diese berücksichtigen zusätzlich zur Aktivität die räumliche Dimension, das heißt die Koordinaten, an denen der ausgelöste Sensor im Käfig platziert ist. Mittels dieser Information kann nun ein Bewegungsprofil jeder Maus erstellt werden, welches helfen kann, die Auswirkungen von Alzheimer auf eine Maus zu untersuchen. Dieses Bewegungsprofil wird bei der ersten vorgestellten Technik für jeden Monat berechnet und anschließend wird jeder Monat einzeln dargestellt.

Während bei dieser Technik die zeitliche Dimension durch mehrere Visualisierungen realisiert wird, bietet die zweite Analysemethode eine einzige Ansicht für die gesamte Lebenszeit der Maus. Hierbei wird das Mäuseleben in drei Abschnitte aufgeteilt, und das für jeden der Abschnitte erstellte Bewegungsprofil wird in eine einzige Visualisierung zusammengefasst. Da der Käfig dreidimensional aufgebaut ist, wurde eine manuelle Dimensionsreduzierung bei beiden Techniken vorgenommen, welche zur einfacheren Darstellung und Interpretierbarkeit der Visualisierungen dient.

4.1 Sensormatrix

Bei dieser Technik wurde versucht, die Bewegungen der Mäuse im Käfig darzustellen. Dazu wurden die Sensoren manuell in eine Reihenfolge gebracht, welche sowohl Abstände als auch den Aufbau des Käfigs zu berücksichtigen versucht. Somit wurden die Sensoren einzelner Kompartimente nicht auseinander gerissen, damit noch Aussagen über das Revierverhalten der Mäuse getroffen werden kann. Die Anordnung der Sensoren wurde anschließend dazu verwendet, eine zweidimensionale Matrix aufzubauen. Mit dieser Matrix können die Bewegungen einer Maus dargestellt werden, wobei die Farbsättigung einer Zelle $M_{i,j}$ aussagt, wie oft die Maus von Sensor i nach Sensor j gelaufen ist. Somit hat die Matrix 27 Spalten und Zeilen (entspricht der Anzahl der Sensoren), wobei die Diagonale der Matrix immer leer ist, da schon im Vorverarbeitungsschritt Mehrfachauslösungen ein und desselben Sensors aus den Daten gelöscht wurden.

Wie schon oben gesagt, werden in den Spalten und in den Zeilen alle 27 Sensoren repräsentiert. Hierfür wird also eine eindimensionale Projektion der Sensoren benötigt, da die Sensoren in der Matrix in eine lineare Reihenfolge gebracht werden müssen. In der Abbildung 4.1 ist der eigentliche Käfig und die gerade vorgestellte Sensormatrix dargestellt. Die Farben sind zum Vergleich des Käfigaufbaus und der vorgestellten Technik eingesetzt worden, wobei gleiche Farben gleiche Bereiche im Käfig repräsentieren. Wie zu sehen ist, werden zusammenhängende Bereiche durch die eindimensionale Projektion nicht voneinander getrennt.

Diese Matrix wird für jeden Monat einzeln erstellt und anschließend werden alle Monate

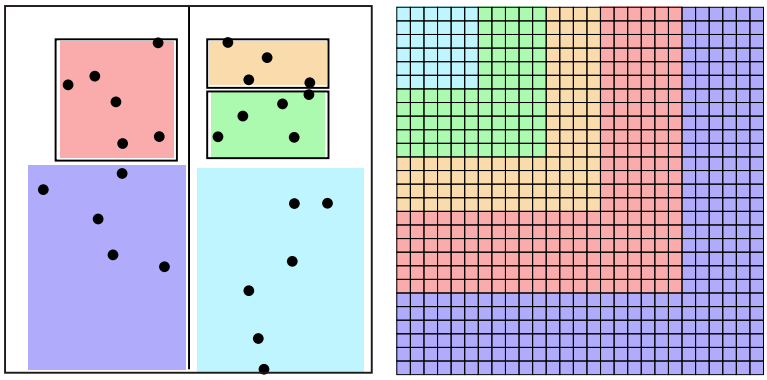


Abbildung 4.1: Hier ist zum Vergleich auf der linken Seite eine zweidimensionale Repräsentation des Käfigs aufgezeichnet. Auf der rechten Seite ist die entsprechende Darstellung mit der Sensormatrix abgebildet, wobei die Farben den korrespondierenden Bereichen im Käfig entsprechen.

nebeneinander angezeigt. Ähnlich der in Abschnitt 3.1.1 vorgestellten Technik zur Untersuchung der Laufstrecken wird auch bei der Sensormatrix ein Differenzbild angeboten, so dass zwei Mäuse einfach miteinander verglichen werden können. Somit ist der Aufbau der Visualisierung wie folgt: Das gesamte Fenster wird in drei Zeilen aufgeteilt. Jede Zeile besteht aus der Schnittmenge der Lebensmonate beider Mäuse, wobei jede Sensormatrix für einen Monat steht. In der ersten Zeile wird das Leben der ersten Maus (blau) und in der zweiten Zeile das Leben der zweiten Maus (rot) dargestellt. Die dritte Zeile bietet einen direkten Vergleich der beiden Mäuse mit dem Differenzbild, das in einer Zelle den Farbton derjenigen Maus annimmt, welche für diese Zelle eine höhere Aktivität hat. Zusätzlich wurden für diese Visualisierung die schon besprochenen Mausprototypen erstellt, wobei wieder der oben beschriebene angepasste Durchschnitt verwendet wurde.

In Abbildung 4.2 wurden männlich gesunde mit männlich kranken Mäusen verglichen. Die Farbsättigung wurde mittels einer logarithmischen Normalisierung berechnet, da die Kontraste beim Druck zu schwach waren. Bei beiden Mausgruppen fällt ein nur leicht unterschiedliches Revierverhalten auf, wobei das Revierverhalten gesunder Mäuse etwas ausgeprägter ist als das der kranken Mäuse. Interessant ist auch die Symmetrie der Sensormatrix, die vom Augenschein her bei beiden Mausgruppen gleich stark ausgeprägt ist. Diese Symmetrie bedeutet eigentlich, dass die Mäuse den Weg von einem zu einem anderen Sensor gleich häufig hin wie zurück laufen. Bei den Verbindungen zwischen zwei voneinander abgetrennten Bereichen ist die Symmetrie nicht erstaunlich, da sonst nur ein Rundweg durch den ganzen Käfig wieder zum Ausgangspunkt führen würde. Viel interessanter ist aber die Symmetrie, die innerhalb eines Bereiches auftritt, wo gar kein Zwang vorherrscht, den gleichen Weg wieder zurück zu gehen. Es kann aber auch sein, dass diese Symmetrie durch die ein Monat umfassende Aggregation der Bewegungen entsteht, bei der sich die gelaufenen Wege einfach nur mitteln. Im Differenzbild erkennt man, dass die Bewegungsintensität der kranken Mäuse im Bezug zu den gesunden Mäusen im Laufe des Käfiglebens abnimmt. Dies ist kein künstlich generiertes Artefakt, da auch für den angezeigten Mausprototyp der angepasste Durchschnitt verwendet

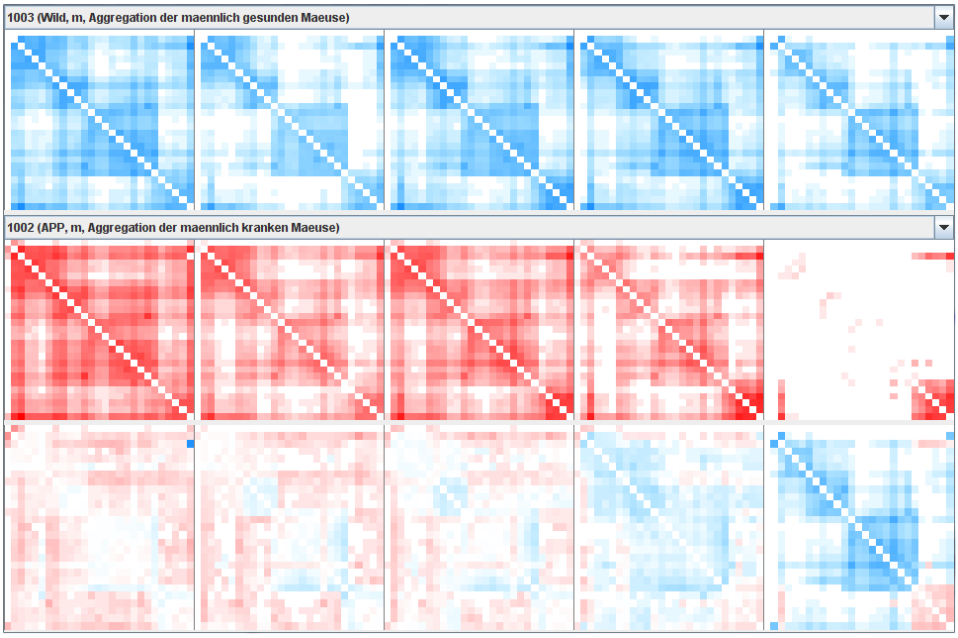


Abbildung 4.2: Dieses Bild zeigt einen Vergleich von männlichen gesunden (blau) mit männlichen kranken Mäusen (rot). Die Farbwerte wurden mittels einer logarithmischen Normalisierung berechnet.

wurde, welcher die zur jeweiligen Zeit im Käfig lebende Population berücksichtigt.

4.2 Northern Lights Map

Die Northern Lights Map soll das vollständige Leben im Käfig sowohl zeitlich als auch räumlich darstellen. Hierzu wird der dreidimensionale Käfig zunächst auf eine zweidimensionale Ansicht projiziert. Diese zweidimensionale Repräsentation ist in Abbildung 4.3 dargestellt, bei der zusätzlich noch die Sensoren mit schwarzen Punkten und die Käfigabtrennungen und -ebenen mittels schwarzer Linien eingezeichnet wurden.

Für die Visualisierung wird das Leben einer Maus auf Monatsbasis in drei Abschnitte aufgeteilt. Diese drei Abschnitte werden benutzt, um Auslösungshäufigkeiten eines Sensors in die drei Farbkanäle zu legen. Somit wird bei dieser Technik ein dreidimensionaler Colormap verwendet, welcher mittels der Farbe die zeitliche Dimension der Mausbewegungen kodiert.

Im Detail werden für jeden dieser drei Abschnitte die Aufenthaltshäufigkeiten an jedem Sensor berechnet und auf die zweidimensionale Käfigrepräsentation als Höheninformation gelegt. Zunächst kann die Höhe nur über den Sensoren ungleich Null sein, an allen anderen Stellen ist sie Null. Da jedoch ein einzelnes Pixel für die Wahrnehmung nicht sehr anschaulich ist und ein farbig ausgefüllter Kreis nicht besonders intuitiv ist, wird ein Tiefpassfilter auf die Höhenkarte angewendet. Der verwendete Filter ist ein Gauß'scher Filter mit großem Filterkernel (in den vorgestellten Beispielen hat σ den Wert 25). Der Gauß'sche Filter wird benutzt, da er einfach zu verwenden und – viel wichtiger – separierbar ist. Durch die Separierbarkeit hat

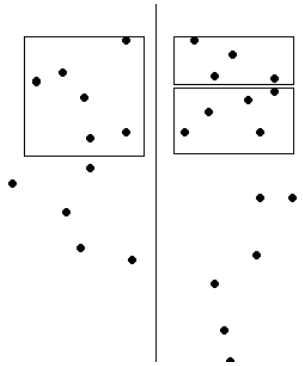


Abbildung 4.3: Diese Abbildung ist die zweidimensionale Repräsentation des Käfigs, wobei die schwarzen Punkte die Sensoren und die schwarzen Linien Abtrennungen bzw. verschiedene Ebenen darstellen.

diese Filterfunktion keine quadratische, sondern trotz einer zweidimensionalen Anwendungsfläche noch immer lineare Laufzeit, was sich bei einer Kartenauflösung von 300×300 Pixeln deutlich bemerkbar macht. Das resultierende gefilterte Höhenfeld – für jeden Zeitabschnitt einzeln erstellt – wird nun auf das Intervall $[0, 1]$ normalisiert. Bei diesem Schritt können nun verschiedene Normalisierungen, wie beispielsweise lineare oder Quadratwurzel Normalisierungen verwendet werden. Diese wirken sich auf die Farbsättigung der Visualisierung aus. Jedes der normalisierten Höhenfelder wird danach in einen Farbkanal gelegt und somit in eine einzige Visualisierung zusammengeführt.

Daher wird mit dieser Technik sowohl die räumliche als auch die zeitliche Information der Daten verwendet. Die zeitliche Dimension der Daten wird mittels des Farbwertes ausgedrückt und die Sättigung der Farben zeigt an, wie oft eine Maus einen bestimmten Sensor besucht hat. In der Abbildung 4.4 wird ein Ausschnitt aus dem verwendeten Colormap gezeigt. Der gezeigte Colormap kann noch über die Sättigung variieren, somit wird hier nur eine Schnittebene eines Prismas abgebildet. Durch das Zusammenführen der drei Höhenfelder in einem einzigen Bild werden in den drei Farbkanälen die unterschiedlichen Lebensabschnitte gespeichert: im Rotkanal wird das erste, im Grünkanal das zweite und im Blaukanal das dritte Lebensdrittel der Maus dargestellt. Gleichmäßige, sich nicht über die Zeit verändernde Aufenthaltshäufigkeiten (die Werte in den drei Farbkanälen sind also alle gleich), werden durch Grautöne angezeigt. Mischfarben resultieren aus zwei hohen Werten und einem niedrigen Wert in den Farbkanälen, Gelb wird beispielsweise aus der Kombination des roten und grünen Farbkanals gebildet.

Da bei dieser Technik die Lebenspannen auf Monatsbasis gedrittelt wurden, mussten die Mäuse mindestens 3 Monate gelebt haben, um geeignet zu sein. Dies war leider bei einigen Mäusen nicht der Fall, so dass diese Mäuse bei der hier vorgestellten Technik nicht berücksichtigt werden konnten. Durch diese Einschränkung wurden zehn von 29 gesunden weiblichen Mäusen, acht von 33 gesunden männlichen Mäusen und drei von 11 kranken männlichen Mäusen nicht untersucht. Die Einschränkung war zwar nicht notwendig, da auch auf Tagesbasis hätte gedrittelt werden können, jedoch bleiben mit noch ca. 75 % der Mäuse genügend übrig, um Aussagen treffen zu können.

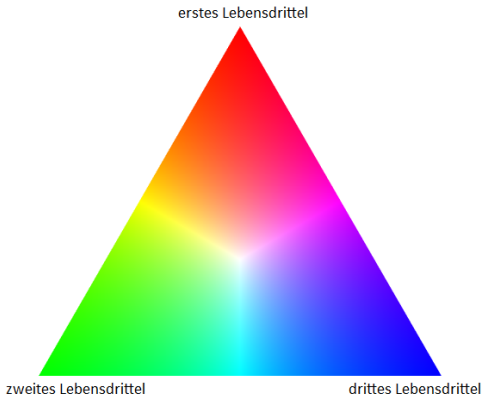


Abbildung 4.4: Diese Abbildung zeigt einen Ausschnitt aus dem verwendeten Colormap, wobei im Rotkanal das erste, im Grünkanal das zweite und im Blaukanal das dritte Lebensdrittel der Maus kodiert wird. Die Farbsättigung gibt die Häufigkeit eines Sensorbesuches an.

In Abbildung 4.5 wird die Northern Lights Map einer männlich gesunden Maus gezeigt, wobei die Farbsättigung mittels einer Quadratwurzelnormierung berechnet wurde. Diese Maus wurde ausgewählt, da bei ihr der Lebensablauf besonders außergewöhnlich ist. Ihr Leben beginnt sie in der linken unteren Ecke des Käfigs und verlegt ihr Revier im Laufe ihres Lebens im Uhrzeigersinn bis sie schließlich am Ende ihres Lebens in der rechten unteren Ecke angekommen ist. Mittels des grünen und hellblauen Rahmens wird noch kodiert, dass die Maus gesund (grün) und männlich (hellblau) ist. Alle 62 für diese Analyse geeigneten Mäuse werden zusammen im Anhang in Kapitel 6 gezeigt.

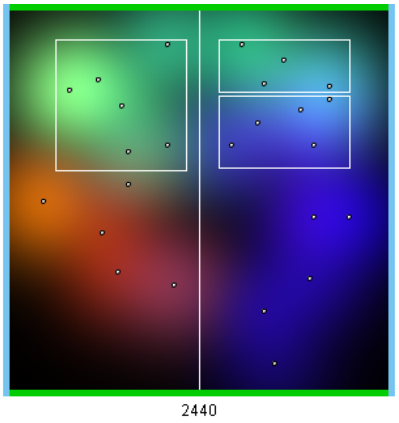


Abbildung 4.5: Hier wird stellvetretend die Northern Lights Map einer männlich gesunden Maus gezeigt. Gut sichtbar ist der im Uhrzeigersinn verlaufende Lebensweg der Maus von der unteren linken Ecke des Käfigs bis hin zur unteren rechten Ecke. Die Farbsättigung wurde hierbei mit einer Quadratwurzelnormierung berechnet.

Zusätzlich wurde auf den so generierten Daten ein Graphlayoutalgorithmus des GraphViz-

Tools [1] angewendet. Dieser Schritt wurde durchgeführt, um zu untersuchen, ob es möglich ist die Mäuse anhand der Sensorbesuche in den Lebensdritteln zu unterscheiden. Hierzu wurde ein 83 - dimensionaler Featurevektor erzeugt, welcher über alle Mäuse auf das Intervall $[0, 1]$ linear normalisiert wurde. Dieser Vektor besteht aus dreimal (alle drei Lebensabschnitte) 27 Sensorbesuchshäufigkeiten und zusätzlich aus dem Durchschnitt und der Varianz aller Häufigkeiten. Die Varianz wird verwendet, um Revierverhalten in die Ähnlichkeit miteinzubeziehen. Die Varianz ist also bei allen Mäusen groß, die nur in einem kleinen Teil des Käfigs aktiv waren. Anschließend wurden die Featurevektoren dazu verwendet eine Distanzmatrix zu berechnen. Bei der Berechnung wurde die Manhattan-Distanz verwendet, da andere Distanzmaße viel weniger Sinn machen. Die Matrix konnte nun in einen Graphen umgewandelt werden, bei dem die Mäuse einzelne Knoten sind und die Distanzen die Längen der Kanten zwischen den Mäusen bzw. Knoten darstellen. Mittels dieser Informationen über den zu zeichnenden Graphen konnte das GraphViz-Tool die Abbildung 4.6 erzeugen. Dieser Layoutalgorithmus führt ein multi-dimensional scaling durch, um eine globale Energiefunktion zu minimieren. Hierbei wird das Optimierungsproblem als eine Menge von geforderten Distanzen zwischen Paaren von Knoten gegeben. Der Algorithmus versucht den Fehler zwischen allen tatsächlichen und geforderten Distanzen zu minimieren. In dieser Abbildung sind die 62 Mäuse als Kreise eingetragen worden, wobei die Farben die Eigenschaften einer Maus beschreiben. Die Füllfarbe gibt an, ob die Maus männlich (blau) oder weiblich (rosa) ist, und die Umrissfarbe zeigt, ob eine Maus krank (rot) oder gesund (grün) ist. Die schwarzen Beschriftungen an den Knoten dienen dazu, die Maus eindeutig zu identifizieren, um einen Vergleich mit den entsprechenden Northern Lights Maps in Kapitel 6 durchführen zu können.

4.3 Ergebnisse

In diesen Kapitel wurden die Daten nicht nur als reine Zeitreihen gesehen, sondern der räumliche Bezug der Ereignisse wurde mitberücksichtigt. So wurde in der ersten vorgestellten Technik die Bewegungsrichtungen einer Maus von einem Sensor zum anderen untersucht und mittels einer Matrix dargestellt. Mittels dieser Methode konnte beispielsweise ein deutlich unterschiedliches Verhalten von weiblichen zu männlichen Mäusen festgestellt werden. Jedoch war es nicht möglich, zwischen kranken und gesunden Mäusen zu unterscheiden. Zusätzlich fiel eine starke Symmetrie der Matrix auf, welche sowohl bei den kranken als auch den gesunden Mäusen auftrat. Diese Symmetrie könnte darin begründet sein, dass einer Maus nur eine begrenzte Anzahl von unterschiedlichen Wegen zur Verfügung steht und sich dadurch die Wege einfach nur mitteln. Bei der Aggregation der Bewegungen durch Mausprototypen konnte noch eine Aufteilung des Käfigs beobachtet werden.

Die Northern Lights Map bietet die Möglichkeit, (fast) alle Mäuse miteinander zu vergleichen, wobei die räumliche Dimension der Daten intuitiv sichtbar ist und die zeitliche Dimension mittels des Farbtons dargestellt wird. Die männlichen Mäuse, die ihr Territorium verteidigen und keine Rivalen dulden, sind in ihrer Bewegungsfreiheit eigentlich eingeschränkt. Dies wird auch in den Northern Lights Maps der männlichen Mäuse sichtbar, da die einzelnen Käfigbereiche sehr unterschiedlich koloriert sind. Die männlichen Mäuse halten sich also zu jeder Zeit nur in einem kleinen Bereich des Käfigs auf. Bei den weiblichen Mäusen ist vorwie-

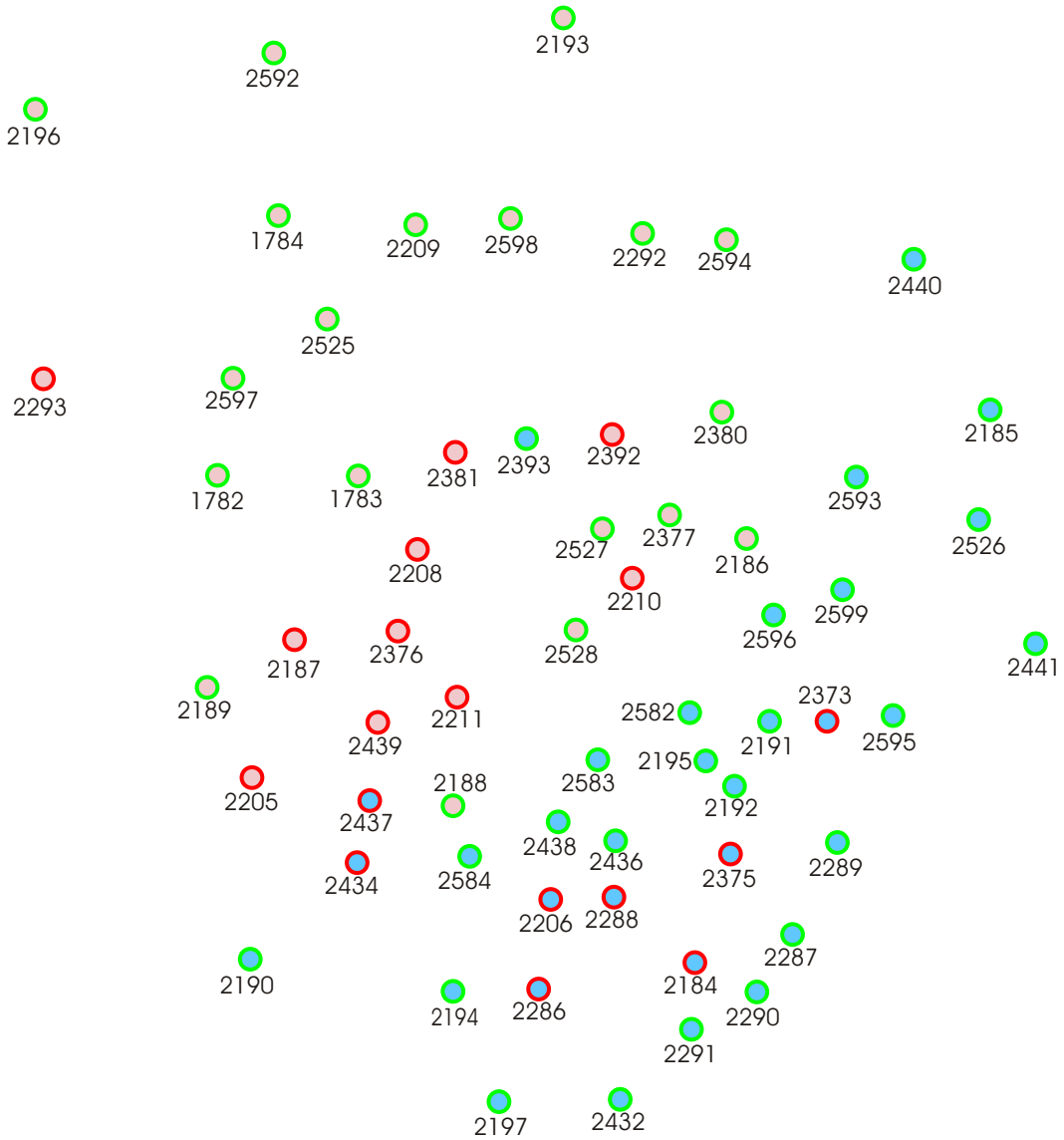


Abbildung 4.6: In dieser Abbildung wird das Ergebnis des Graphlayoutalgorithmus gezeigt. Jeder Kreis steht für eine Maus und die Füllfarbe gibt an, ob die Maus männlich (blau) oder weiblich (rosa) ist. Mittels der Umrissfarbe wird angezeigt, ob eine Maus krank (rot) oder gesund (grün) ist.

gend der ganze Käfig in einem gleichen, pastellfarbenen Farbton gefärbt, wobei die kranken weiblichen Mäuse eine stärker farbige Northern Lights Map als die gesunden weiblichen Mäuse haben. Außerdem fällt auf, dass fast alle männliche Mäuse über ihr Leben lang nicht das gleiche Revier bewohnen. Besonders gut sichtbar wird diese Tatsache an der männlichen gesunden Maus 2440, da diese im Laufe ihres Lebens den ganzen Käfig durchwandert. Dies liegt entweder daran, dass die Mäuse aus ihren Revieren vertrieben werden, oder es ihnen zu langweilig ist, ihr ganzen Leben in einem Teilbereich des Käfigs zu verbringen. Interessant wäre hierbei zu wissen, ob die dominanten Tiere ihr Revier nie verlagern oder ob die unterlegenen Tiere an einen Teilbereich des Käfigs gebunden sind.

Die Anordnung der Mäuse mittels einem Graphlayoutalgorithmus auf Grund der Ähnlichkeit der den Northern Lights Maps zugrunde liegenden Daten zeigt, was überhaupt an Information in den Daten steckt. Besonders wichtig bei der Featurevektorerstellung war es, die Varianz als eigene Dimension einfließen zu lassen. Ohne diese zusätzliche Dimension gelang keine gute Aufteilung der Mäuse, da ohne Varianz das Revierverhalten der Mäuse nicht berücksichtigt wurde. Auch war die lineare Normalisierung der Daten sehr wichtig, um eine Vergleichbarkeit und somit sinnvolle Anordnung der Mäuse zu erhalten. Diese Erzeugung der Featurevektoren sorgte für eine gute Unterscheidung der Geschlechter. Schließlich kann eine einfache Trennlinie durch den Graphen gelegt werden, um fast alle Mäuse geschlechterspezifisch voneinander zu trennen. Die Aufteilung der Mäuse auf Grund ihres Gesundheitszustandes ist jedoch schwieriger. So liegen die kranken Mäuse zwar nicht zufällig gestreut unter den gesunden Mäusen, aber auch nicht richtig gut als eine Gruppe vor. Anhand der Lage der kranken Mäuse kann man davon ausgehen, dass sich die kranken männlichen Mäuse kaum von ihren gesunden Geschlechtsgenossen unterscheiden. Jedoch fällt auf, dass die kranken weiblichen Mäuse in der Nähe der männlichen Mäuse angeordnet sind. Dies lässt darauf schließen, dass die Auswirkungen von Alzheimer auf weibliche Mäuse stärker sind als auf ihre männlichen Artgenossen. Ferner ist der Abstand zwischen den männlichen Mäusen geringer als der zwischen den weiblichen Mäusen. Anscheinend verhalten sich männliche Mäuse bei der Häufigkeit der Sensorbesuche ähnlicher als weibliche Mäuse.

5 Zusammenfassung und Ausblick

In dieser Arbeit wurden einige Analysemethoden zur Untersuchung von aufgezeichneten Mausbewegungen in einem Käfig vorgestellt. Zuerst wurde versucht, ohne den räumlichen Bezug der Daten zu beachten, eine Unterscheidung der Mausgruppen vorzunehmen und anschließend wurde im zweiten Teil dieser Arbeit die räumliche Dimension der Mausbewegungen mit berücksichtigt.

Durch das Nichtberücksichtigen der räumlichen Dimension mussten Maße zur Untersuchung von Bewegung und Trinkhäufigkeit aufgestellt werden. Diese dienten der Analyse der RFID - Aufzeichnungen. So wurde beispielsweise die gelaufene Strecke oder das Verhältnis von Trinkhäufigkeit zur gelaufenen Strecke einer Maus untersucht.

Diese neuen Maße wurden anschließend mit Recursive Patterns visualisiert. Hierbei wurde eine Erweiterung des Recursive Pattern zum besseren Vergleich zweier Mäuse bzw. Mausprototypen vorgestellt. Mittels eines Differenzbildes und unter Verwendung eines intuitiven Colormaps kann sofort verglichen werden, welche Maus einen höheren oder niedrigeren Wert hat.

Die statistische Auswertung des Verhältnisses von Trinkhäufigkeit zur zurückgelegten Streckenlänge ergab signifikante Unterschiede zwischen den gesunden und kranken Mäusen. Außerdem konnte gezeigt werden, dass sich weibliche Mäuse durch Alzheimer in ihrem Trinkverhalten stärker verändern als männliche Mäuse.

Zur Visualisierung der Daten mit räumlichem Bezug wurden zwei Techniken vorgeführt. Bei der ersten Technik wurden die Mausbewegungen mit einer Sensormatrix dargestellt. Mittels dieser konnte das Revierverhalten von zwei Mäusen oder Mausprototypen miteinander verglichen werden. Anschließend wurde die Northern Lights Map Technik vorgestellt, welche eine farbige Repräsentation der Aufenthaltshäufigkeiten bietet. Hierbei wurde durch das Dritteln der Lebensspanne einer Maus und das Ausnutzen der drei Farbkanäle des Monitors eine intuitive Visualisierung geschaffen.

Die für die Northern Lights Maps erzeugten Daten wurden zuletzt noch für eine Anordnung der Mäuse als Knoten eines Graphen verwendet. Dieser Schritt diente zum Aufzeigen der möglichen Separierbarkeit der Daten unter Berücksichtigung von Raum und Zeit. Die Unterscheidung von männlichen zu weiblichen Mäusen fiel dabei besser aus als die Unterscheidung von kranken und gesunden Mäusen.

Insgesamt konnte ein unterschiedliches Bewegungsmuster und unterschiedliche Verhaltensweise von kranken / gesunden, männlichen / weiblichen Mäusen gefunden werden. Somit war die Untersuchung der Daten mittels der vorgestellten Techniken erfolgreich und führte zu interessanten Einblicken in das Leben einer Maus.

Zukünftige Arbeit könnte das Erweitern der Northern Lights Map zur Aufteilung der Daten auf Tagesbasis sein, damit alle Mäuse trotz ihres manchmal zu kurzen Lebens berücksichtigt werden. Außerdem könnte versucht werden, eine automatische Unterscheidung von kranken

und gesunden Mäusen auf Grund der gewonnen Erkenntnisse zu erreichen. Auf jeden Fall erfolgversprechend sollte die automatische Klassifikation von männlichen und weiblichen Mäusen sein, da sich diese Mausgruppen stark voneinander unterscheiden.

6 Anhang

Die folgenden Abbildungen zeigen die Northern Lights Maps aller 62 Mäuse, die für dieses Verfahren geeignet sind. Die senkrechten Rahmenlinien für das Geschlecht, wobei rosa eine weibliche und hellblau eine männliche Maus ist. Die waagerechten Rahmenlinien geben den Gesundheitszustand einer Maus an. Mit roten Linien wird eine kranke Maus und mit grünen Linien eine gesunde Maus angezeigt.

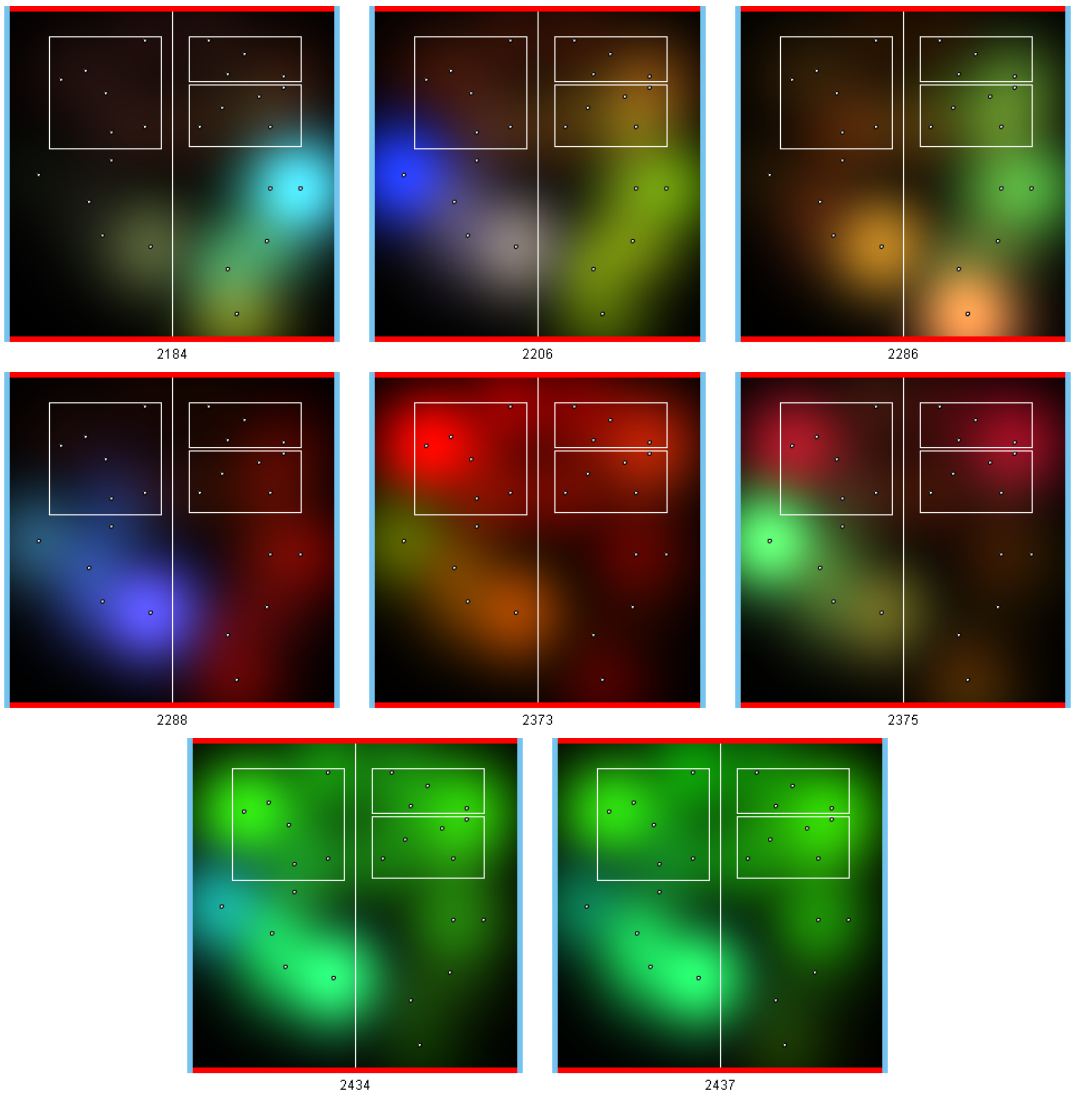
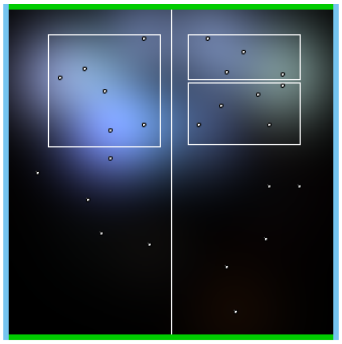
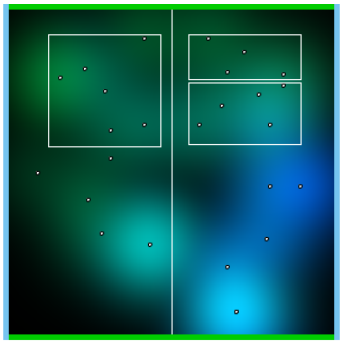
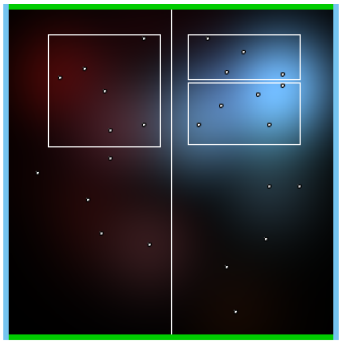
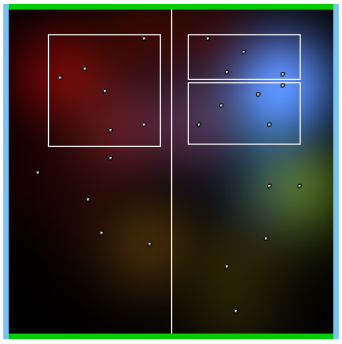


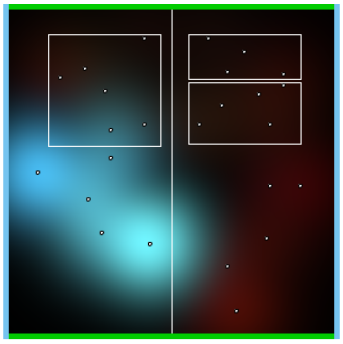
Abbildung 6.1: Northern Lights Maps aller männlichen kranken Mäuse

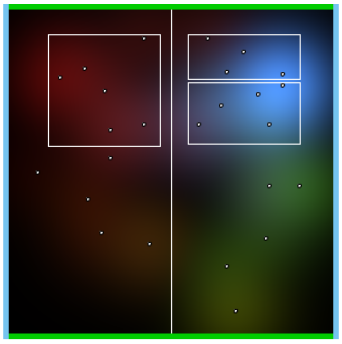

2185

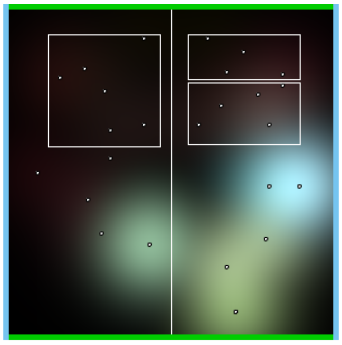

2190

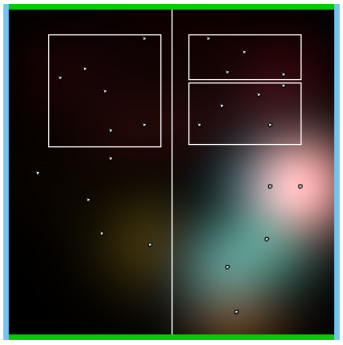

2191

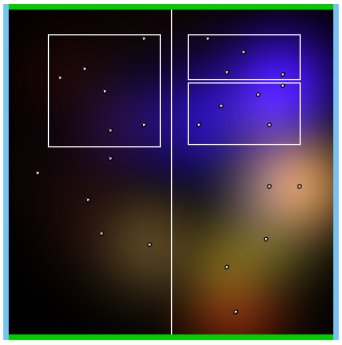

2192

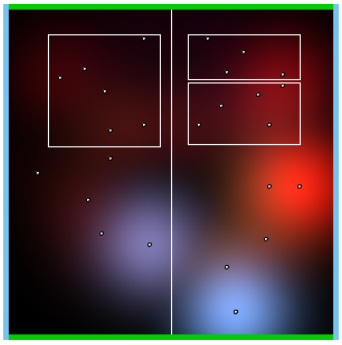

2194

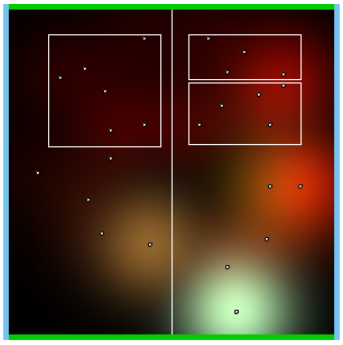

2195

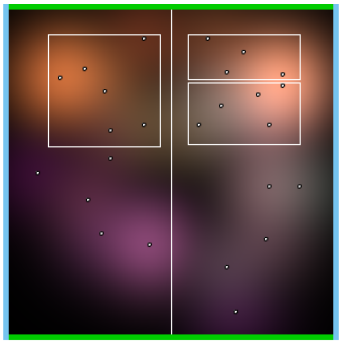

2197

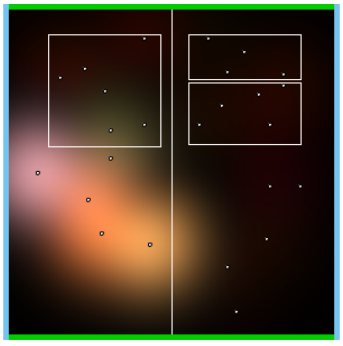

2287


2289


2290


2291


2393


2432

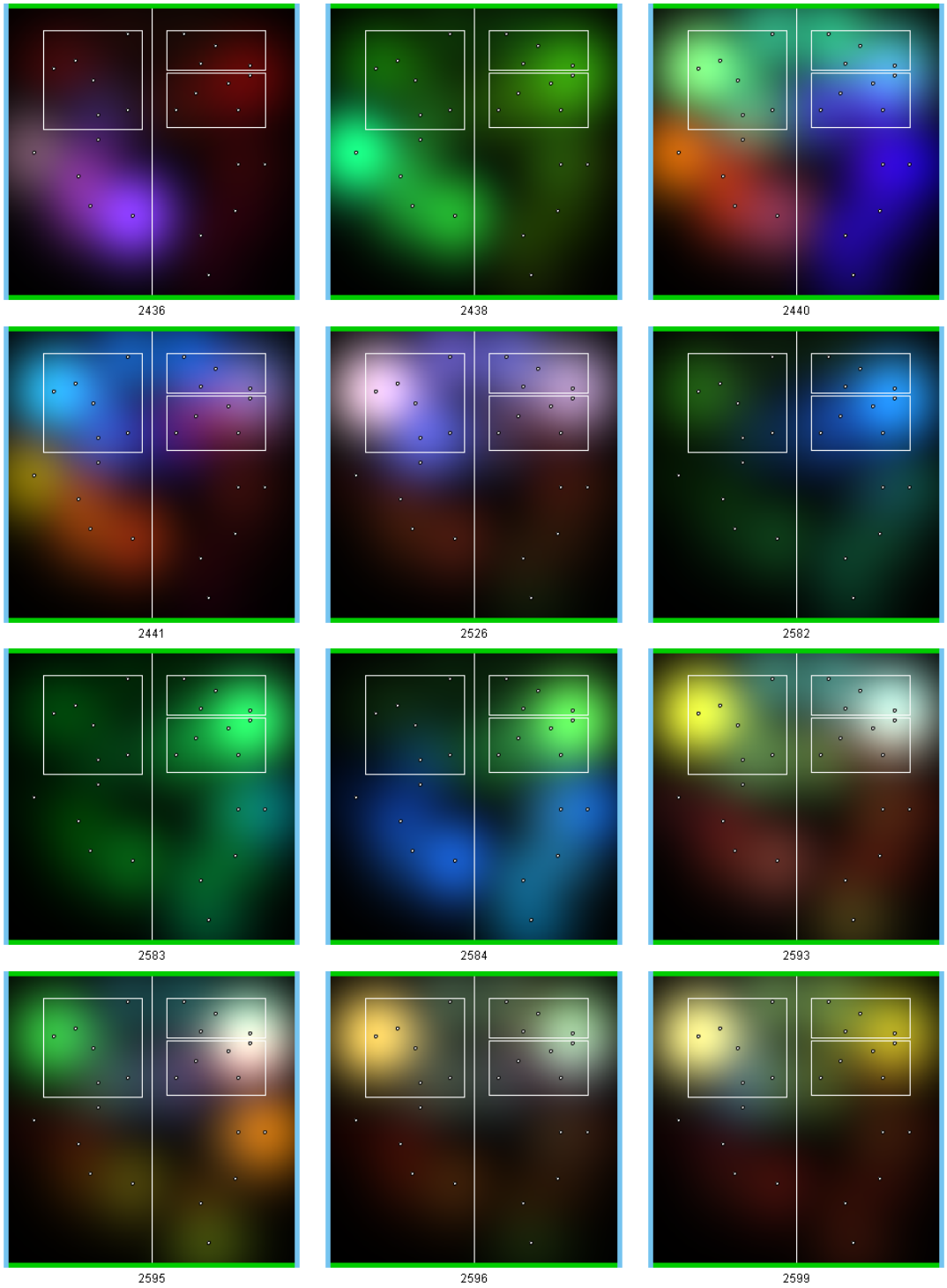


Abbildung 6.2: Northern Lights Maps aller männlichen gesunden Mäuse

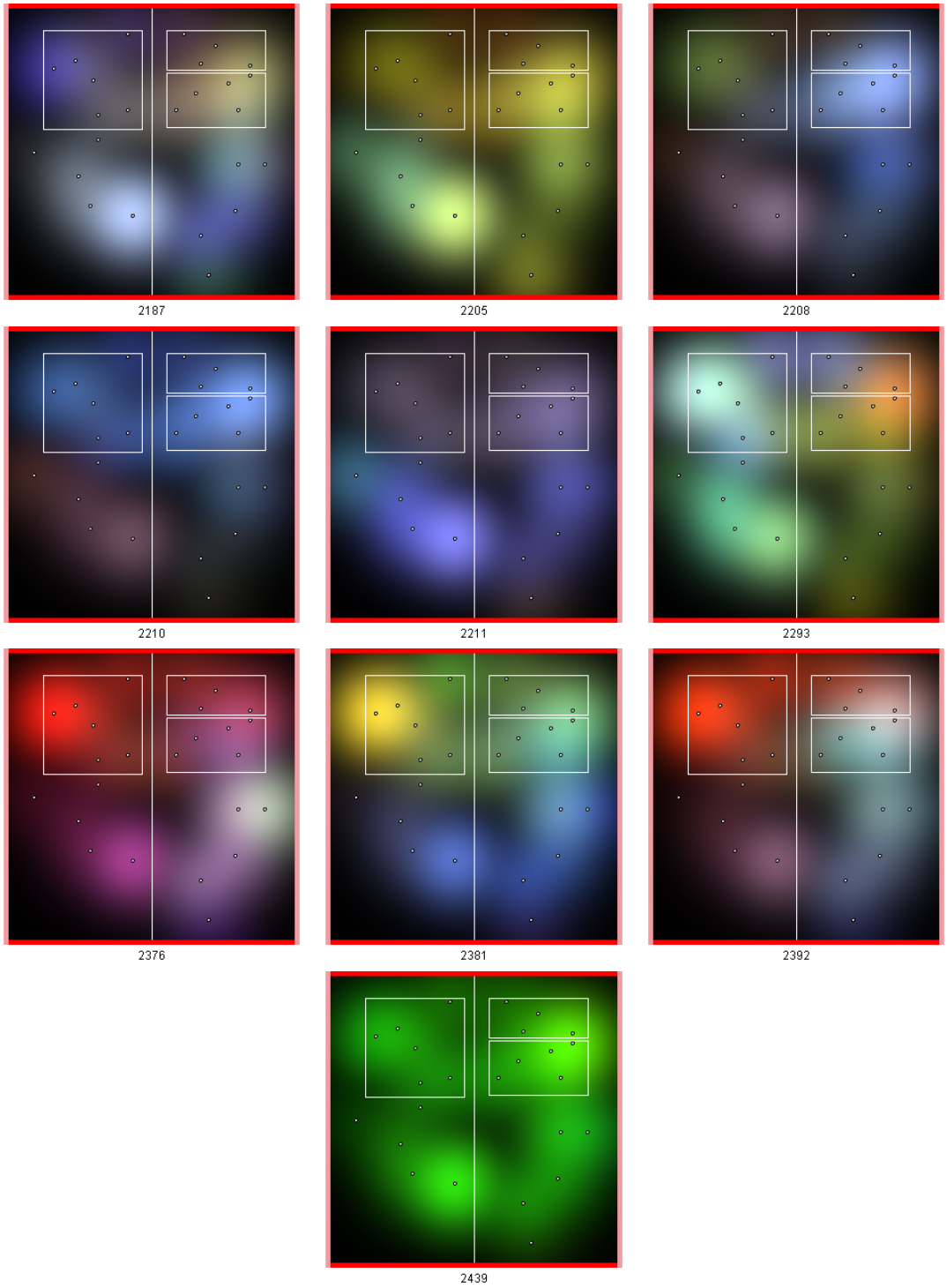
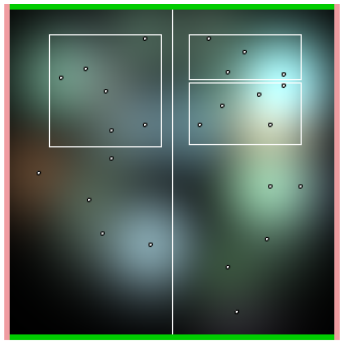
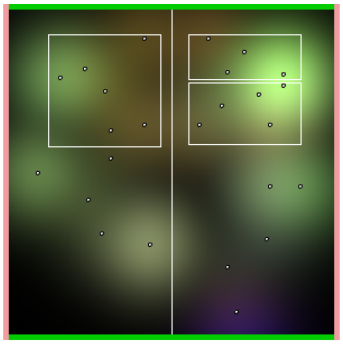
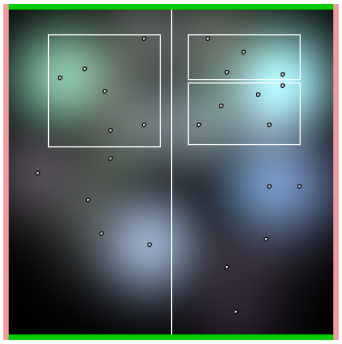
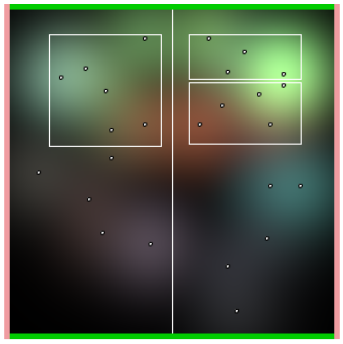


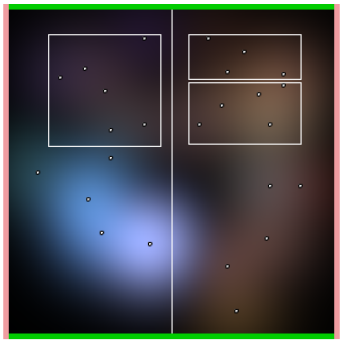
Abbildung 6.3: Northern Lights Maps aller weiblichen kranken Mäuse

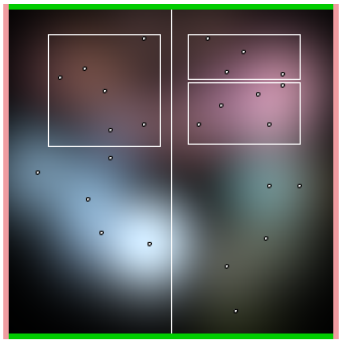

1782

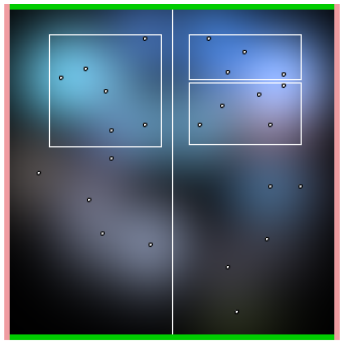

1783

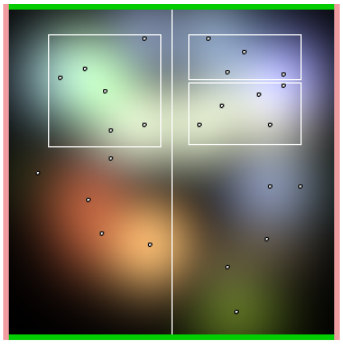

1784

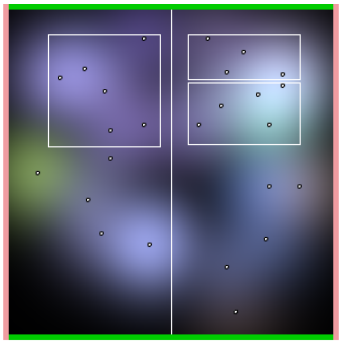

2186

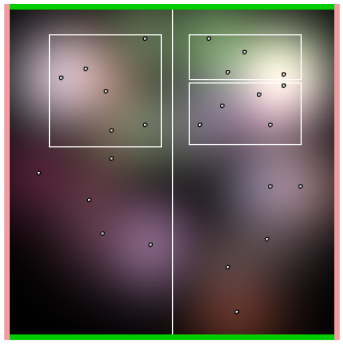

2188

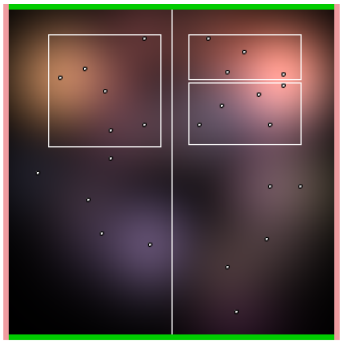

2189

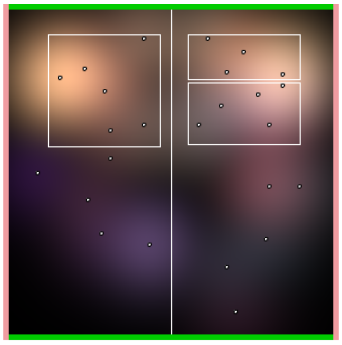

2193

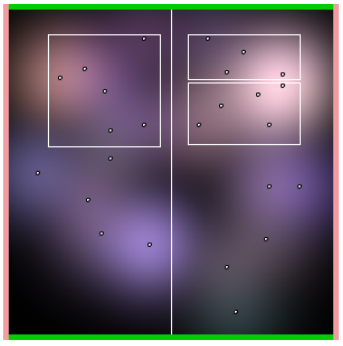

2196

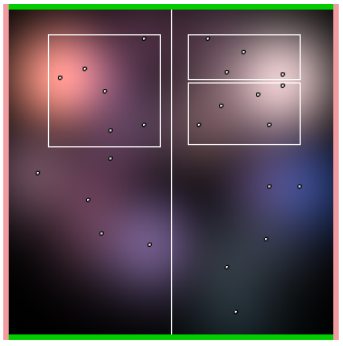

2209


2292


2377


2380


2525


2527

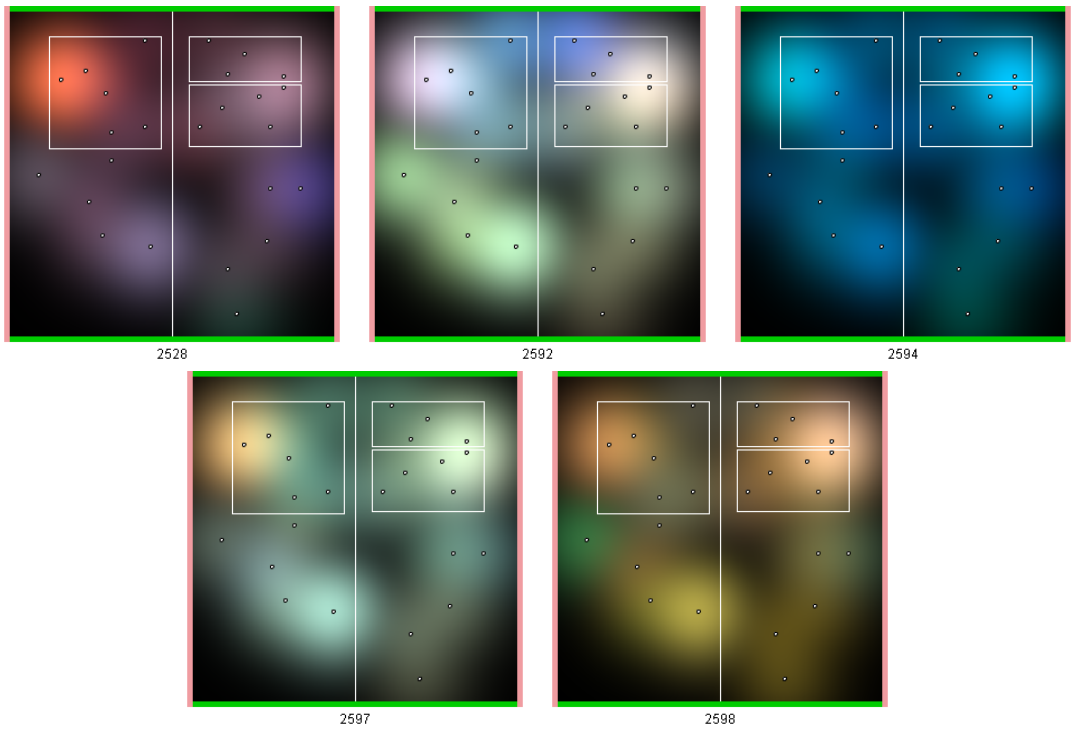


Abbildung 6.4: Northern Lights Maps aller weiblichen gesunden Mäuse

7 Danksagung

An allererster Stelle möchte ich meinem Betreuer Herrn Dr. Florian Mansmann für seine hilfreichen Hinweise und Anmerkungen danken. Ohne sie wäre diese Bachelorarbeit in der jetzigen Form nicht denkbar gewesen. Außerdem gebührt noch Frau Mareike Kritzler von der Universität Münster Dank, da sie die Arbeit mit den Mausdaten erst durch ihre Erläuterungen durchführbar machte. Ferner möchte ich noch Herrn Prof. Dr. Daniel A. Keim meinen Dank dafür aussprechen, dass er mir die Möglichkeit gab, an seinem Lehrstuhl interessante und spannende Projekte und Arbeiten durchzuführen.

Literaturverzeichnis

[1] Graphviz – graph visualization software. http://www.graphviz.org/, 2008.

[2] Anke Altintop. Using color effectively in visualization. (Colormap-Tool: http://infovis.uni-konstanz.de/tools/colormap/index.html), 2002.

[3] F. Brian and J. Pritchard. Visualisation of historical events using lexis pencils. In *Case Studies of Visualization in the Social Sciences*, 1997.

[4] John V. Carlis and Joseph A. Konstan. Interactive visualization of serial periodic data. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 29–38. ACM, 1998.

[5] R Foundation. R – statistical computing tool. http://www.r-project.org/, 2008.

[6] L. Gygax, G. Neisen, and H. Bollhalder. Accuracy and validation of a radar-based automatic local position measurement system for tracking dairy cows in free-stall barns. In *Computers and electronics in agriculture*, pages 22–33, 2007.

[7] J. Hartung, B. Elpelt, and K.-H. Klösener. *Statistik – Lehr- und Handbuch der angewandten Statistik*. Oldenbourg Verlag, 8. Auflage 1991.

[8] Y. Ivanov, C. Wren, A. Sorokin, and I.Kaur. Visualizing the history of living spaces. In *IEEE Transactions on Visualization and Computer Graphics*, pages 1153–1160, 2007.

[9] T. Kapler and W. Wright. Geotime information visualization. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, pages 25–32, 2004.

[10] Daniel A. Keim, Mihael Ankerst, and Hans-Peter Kriegel. Recursive pattern: A technique for visualizing very large amounts of data. In *VIS '95: Proceedings of the 6th conference on Visualization '95*, page 279. IEEE Computer Society, 1995.

[11] M. J. Kraak. The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference*, 1995, 1988.

[12] M. Kritzler, L. Lewejohann, and A. Krüger. Analysing movement and behavioural patterns of laboratory mice in a semi natural environment based on data collected via rfid-technology. In *Behaviour Monitoring and Interpretation*, pages 17–28, 2007.

[13] M. Kritzler, L. Lewejohann, A. Krüger, M. Raubal, and N. Sachser. An rfid-based tracking system for laboratory mice in a semi natural environment. In *Pervasive 2006 Workshop „Pervasive Technology Applied - Real-World Experiences with RFID and Sensor Networks“*, 2006.

[14] Jock Mackinlay. Automating the design of graphical presentations of relational information. volume 5, pages 110–141, New York, NY, USA, 1986. ACM.

[15] Robert Spence. *Information Visualization – Design for Interaction*. Pearson Education Ltd., Second Edition 2007.

[16] Christian Tominski, James Abello, and Heidrun Schumann. Axes-based visualizations with radial layouts. In *SAC '04: Proceedings of the 2004 ACM symposium on Applied computing*, pages 1242–1247, 2004.

[17] H. Ziegler, T. Nietzschmann, and D. Keim. Visual exploration and discovery of atypical behavior in financial time series data using two-dimensional colormaps. *11th International Conference Information Visualization (IV '07)*, pages 308–315, 2007.