Interactive Pattern Analysis of Multiple T-Maze Data

Fabrizia Bechtold[1], Hrvoje Abraham[2], Rainer Splechtna[1], and Krešimir Matković[1]

[1]VRVis Research Center in Vienna, Austria
[2]AVL-AST d.o.o, Zagreb, Croatia

Abstract

The Multiple T-Maze study is one of the standard methods used in ethology and behaviourism. In this paper we extend the current state of the art in analysis of Multiple T-Maze data for animal cohorts. We focus on pattern finding within animals' paths. We introduce the Sequence View which makes it possible to quickly spot patterns and to search for specific sub-paths in animal paths. Further, we also evaluate four different metrics for string comparison and two widely used embeddings to support interactive clustering. All views are fully integrated in a coordinated multiple views system and support active brushing. This research represents a step towards (semi)-automatic clustering for Multiple T-Maze cohort data, which will significantly improve the Multiple T-Maze data analysis.

1. Introduction

Ethology and behaviourism deal with the scientific and objective study of animal behaviour. Such studies are not performed only to understand animals, but also to better understand physiological processes, neural mechanisms, or learning and aging processes in humans. Conducting control studies with humans is almost technically impossible, and clearly ethically unacceptable. Further, aging processes last very long in humans, and rodents, for example, have a much shorter expected life time and a faster aging (the expected lifetime of Sprague-Dawley rats is about 2 years [Bir13]). A rodent's brain shows significant similarities with the human brain [BN06], which makes rodents a premium mean for studying various processes. A better understanding of learning and memory in relation to the aging of rodents leads to a better understanding of similar processes of humans [BLJ94, HES*06, WPF09, LWF11]. Understanding of the aging process leads to a better health care and helps to cope with the problems of today's aging society. The idea of studying the rodents' behavior and their learning process is not new. Willard S. Small used the behavior of rats in mazes as a measure of learning already at the begin of the 20th century [Sma01, SN27].

Various types of ethological studies have been designed for different research tasks. The Multiple T-Maze, a maze composed of multiple T-shaped segments, is a widely used experiment for studying learning processes. The maze is built so that at each junction the view to the left and to the right looks exactly the same. There are no visual clues on the correct way. As a motivation there is always a reward (food) at the end. Animals are set in the maze several times a day during the first week, and once more after another week. The animals in the maze have to memorize the path using short term (daily repetitions) and long term memory (final repetition after one week). One method to analyze the animals behaviour is to track their movements throughout the maze and compute trajectories.

Evaluation of these trajectories is done using quite a simple statistical evaluation. Bechtold et al. [BSM18] introduced an interactive visual approach to the analysis of Multiple T-Maze data. Our work extends such an approach by means of an automatic analysis. The experts are interested in trajectories of the cohort of animals. They want to see if there are patterns in the behavior of the animals. Are there certain sub-paths that appear more often than some others? Are there clusters of the paths? Answering such questions with a pure interactive solution is tedious. Automatic analysis should complement an interactive solution. However, the automatic analysis of such complex data is challenging. Due to data complexity, the analysis requires a visual feedback and means to refine results interactively. Only a combination of interactive and automatic system can yield satisfactory results here.

In this paper we focus on the animal trajectories in a Multiple T-Maze. Our results should be seen as an extension of an interactive approach as suggested by Bechtold et al. [BSM18], and not as an alternative to interactive analysis methods. In order to explore trajectories, we also use sequence search mechanisms, where users can search in the gate sequences of the animal cohort. We implemented the newly proposed approach into a coordinated multiple views system which is used to provide feed-back, refine searches, and define cohort subsets. The main contribution of this paper is a novel way of analysis for Multiple T-Maze data which combines interactive and automatic analysis methods. In addition, we also describe in detail how we analyze Multiple T-Maze data, and integrate everything in an interactive visualization system.

2. Related Work

Our work is related to several research directions as our approach involves an interactive visual analysis system which takes advantage of humans in the exploration loop, and of automatic analysis methods. Such combinations belong to the field of visual analytics [TC05, KKS*11], whose systems have been deployed in numerous domains. In addition, we rely on the well known coordinated multiple views paradigm [Rob07].

Probably the most related papers to ours is previous work on interactive visual analysis for open field data [MWSB12] and for Multiple T-Maze data [BSM18]. In contrast to those papers, we use different metrics for sequences data and we then project the sequences to a 2D embedding using multidimensional scaling [CC00] and t-SNE methods [vdMH08]. Additionally we incorporate sequence search mechanisms.

Visual analysis of various movement data is a well researched topic [AAB*13]. There are also papers dealing with more specific movement tasks, e.g., movement in sports [SJL*18], movement in air traffic [AAFG18], or movement of animals in the wild [SvL16]. Adrienko et al. [AA13] examine visualisation techniques and movement clustering on the example of primarily traffic trajectories. He et al. [HCC*19] show state-of-the-art visualisation techniques for trajectories such as space-time density views and flow maps.

Another application area for way finding tasks is psychology. Siegel and White [SW75], Montelo et al. [MHRW04], and Ji-Sun et al. [KGMQ08], for example, focus on human way finding. If a way finding study has to be conducted in a controlled environment, rodents represent the first choice. The Multiple T-Maze is a well defined and described experiment [HEB*00, PJCK03]. Bubna-Littitz et al. [BLHKN81] found out that the Multiple T-Maze is the best means for studying the learning process as variations appear much earlier than in other methods.

In addition to the use of simple scalar features, such as *total way traveled* or *time needed to find the goal*, we propose to examine paths in more detail. Clearly, it is important if an animal finds the way and how long it takes. But it is also interesting to study the ways of the cohort; are there any patterns in behavior or do animals which do not find the goal get lost in the same way, for example. Our research makes such an analysis possible.

3. Multiple T-Maze Data and Analysis Tasks

The analysis tasks as identified by Bechtold et al. [BSM18] are unsatisfactory for finding patterns within the paths. In this work we expand the analysis tasks to accommodate the experts' wish to detect patterns in the animals' movement behaviour. Experts seek to examine the data of animals following a certain succession of gates, i.e. the sequence of T-segments in which they traverse the maze. Additionally they are interested in how much time animals spent in each successive T-segment. Lastly, as finding similarities between paths can be costly, experts desire a way to (semi-)automatically cluster paths together. The examination of patterns should be possible on a high level, i.e. on the whole cohort, and on a specific group of animals; for example paths containing a specific traversing sequence of the gates or all paths belonging to a specific cluster.
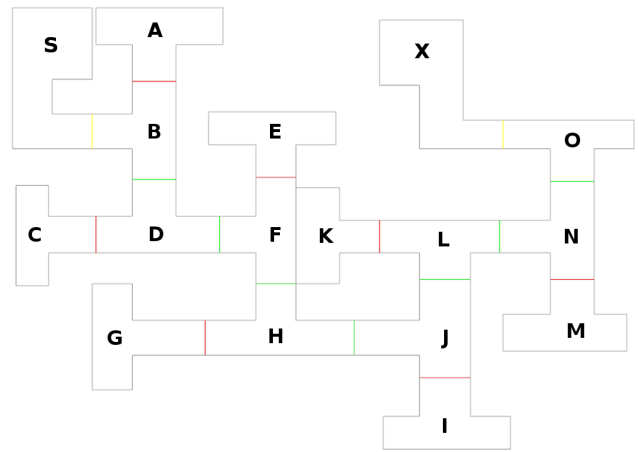

Figure 1: *The Multiple T-Maze is divided into T-shaped segments. Each segment is labeled with a unique letter to enable the computation of sequences. The ideal sequence is SBDFHJLNOX.*

Therefore the tasks are abstracted as follows:

(i) **High Level Tasks**

    H1 Identify sequence of T-segment traverse of all paths
    H2 Identify time spent in each T-segment while traversing
    H3 Cluster similar groups of paths

(ii) **Medium Level Tasks**

    M1 Identify order of traverse of a sub-group of paths
    M2 Analyse paths containing a specific traverse order

The Multiple T-Maze data set consists of various types of attributes. For the computation of descriptive statistics mostly categorical and numerical attributes are used and explored using common statistical views. The paths' trajectories consist of Cartesian coordinates and are the starting point for tackling the above declared tasks. Each gate segment is labeled with a unique letter. Figure 1 shows the used Multiple T-Maze divided into T-segments and the assigned labels. We know in which segment a trajectory coordinate is situated and how many coordinate points are tracked per second. Based on this knowledge we build a data structure combining the label of the traversed segment and time spent in it. For example the value {S, 1.5} means that an animal spent 1.5 seconds in the start area. A sequence data set consists of all successive label and time values throughout a path. To examine the sequences we implemented a new view, the *Sequence View*, which is explained in detail in section 4. The metrics we use for the clustering task **H3** use a string as input data. Therefore we extracted a single sequence string from the trajectories. The string 'SBABD' e.g., represents an animal which moved from the start segment S to the first T-segment B, chose the wrong corridor A and moved back to B, and then to the second T-segment D. The ideal path equates to the string 'SBDFHJLNOX'.

4. Multiple T-Maze Sequences Analysis

The Sequence View is developed to display the animals' traversing sequence in a compact and easy to interact manner. It allows the

visual exploration of all sequences in the data set, as required for task **H1**. A sequence is depicted as a train of boxes, where each box represents a single T-segment or label. To visibly distinguish the individual segments, each label is assigned a color, spanning a gradient from dark brown for the first label, to light blue in the middle and dark blue at the end. This linear color scheme is mapped to the labels in alphabetical order and focuses on the orderly traversal through the maze. Other color mappings could highlight, for example, the deviation from the ideal path. Optionally a legend visually connecting the colors to the segment labels can be displayed. The sequence view shows all sequences at once to give an overview of their structure. To view sequences in more detail, scaling in both directions is possible. Figure 2a shows a vertically scaled section of the Sequence View. This view enables the easy distinguishing of the lengths of the different sequences, indicating how much an animal is moving within the gate. Here it is possible to quickly perceive patterns, such as, for example, a periodic, mirrored pattern of brown, light and dark blue in the 3rd sequence from the top in figure 2a. The sequence depicts repeated movement between the earlier and the later segments, indicating turns along the path.

Optionally the user can select time-dependent sequences. Here the size of the box is not uniformly scaled but it depends on the time spent in the corresponding segment as requested for task **H2**. Figure 2b shows the same section of sequences as figure 2a, but this time with time-dependent scaling. Now it becomes apparent in which segments an animal dwelled for some time and which segments it quickly passed. Comparing these two views, it is easy to see that a sequence with many values does not automatically correspond to a long run. The 7th sequence, for example, is of average length when uniquely scaled, but the longest when time-dependent. I.e., the animal did not move between the segments so much but a lot within single segments. The 7th sequence also has a large dark blue strip, indicating that it did not move out of the last gate segment for some time.

The Sequence View, just as the other views, is embedded into a Coordinated Multiple View system. This allows the selection of a subset of the data set in any view via brushing and highlighting the selection in other views (the requirement for **M1**). The Sequence View is used when sequences themselves are of main interest. For a deeper exploration of the time component in the data, other views are more suitable, as Bechtold et al. [BSM18] have done in their work.

Another important aspect of our work is the brushing of data which follow a specific sequence order, as desired in task **M2**. Therefore we implemented a simple string-based search, as we already have a sequence represented as a single string. This allows to easily select all paths that, e.g., return from segment D to segment B by searching for 'DB'. Though this is possible in the Gate-O-Gon as introduced by Bechtold et al. [BSM18], the other way around, i.e. 'BD', is not easily accomplished. With the string-based search it is also easy to search for paths turning around through specific segments at some point by search for 'BD*DB' for example. It is also possible to exclude a certain sequence, e.g., '~BS' excludes all paths with returns from segment B to the start area. This allows for a far more in-depth exploration of the Multiple T-Maze data.

We focus on sequences themselves, as we want to explore the

data deeply. We are interested if there are some patterns or clusters of sequences which would manifest similar trajectories. The Multiple T-Maze trajectories can be very different. The ideal case is if an animal goes straight to the end area. So, we do have an ideal path. The animals that reach the end area can do it in numerous ways. Exploration of this variation is the main goal of our approach.

In order to cluster trajectories and to see the patterns we need a metric to compare individual trajectories. We have tested four methods known from text sequence comparisons. As the length of a sequence is not necessarily the main criterion for similarity in our case, we choose metrics, which do not penalize difference in length. Imagine an animal that swings between two gates for some time, and then, finally, continues to the end. The swinging would make the sequence very long. Another animal which makes a wrong decision at four different gates (but does it only once per gate) could have a relatively short sequence. The question is now which one is closer to the ideal? Basically there is no right answer to this question. We choose the following four metrics for string comparisons in our study, each using different algorithms to compute the distances to the ideal string and therefore revealing different clusters:

1. Jaro-Winkler
2. longest substring similarity
3. MLIPNS
4. Needleman-Wunsch

The methods were selected by the criteria of assumed capability to determine relative similarity between the routes, i.e., animal behavior exhibiting similar patterns, by calculating the correlation of measured string distance and percentage of common substrings. The chosen methods compensate for different string sizes and have a better correlation with the ratio of common substrings. Additionally, Needleman-Wunsch takes many possible random error patterns into account, which in the case of our animal trajectory data can be interpreted as similar movement behaviour. The Python text distance package [Str] is used to compute the distances. Once the distances are computed they are used to compute 2D embedding of the sequences. We compute MDS – multidimensional scaling [CC00] and t-SNE – t-Distributed Stochastic Neighbor Embedding [vdMH08] for each of the four metrics.

The user can show the embeddings now in a coordinated multiple view setup, and explore the sequences by means of interaction. Figure 3 shows results for the 4 metrics and two embeddings. By viewing the 8 scatter-plots side-by-side the user can detect diverse clusters and explore how they relate to other clusters. The top row shows the MDS and the middle row shows t-SNE approach. Four differently colored brushes are shown in order to illustrate differences between the approaches. The bottom row shows the trajectories views corresponding to different brushes. We scale the point size according to the number of trajectories which are represented by a point. Some algorithms map the same sequences to the same point (which becomes larger then), and some scatter the points.

Note the completely different layouts for the eight cases we examine in order to see which approach could be used for Multiple T-Maze data. We start by brushing the large point in the third view in the top row. It corresponds to ten exactly same sequences. These are ideal paths, as shown in the first bottom view. Some approaches scatter the points quite a lot (see orange points in all scatter plots).
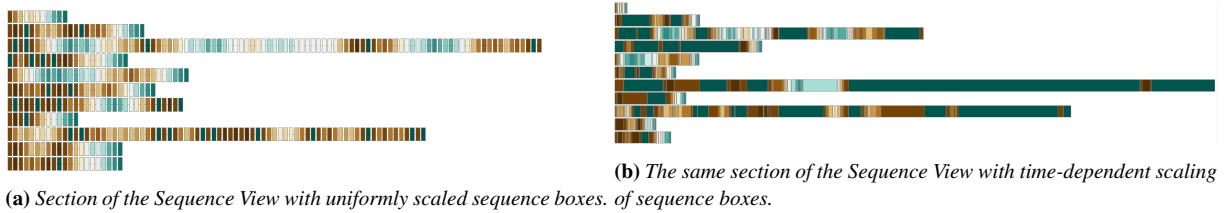
(a) *Section of the Sequence View with uniformly scaled sequence boxes.*

(b) *The same section of the Sequence View with time-dependent scaling of sequence boxes.*
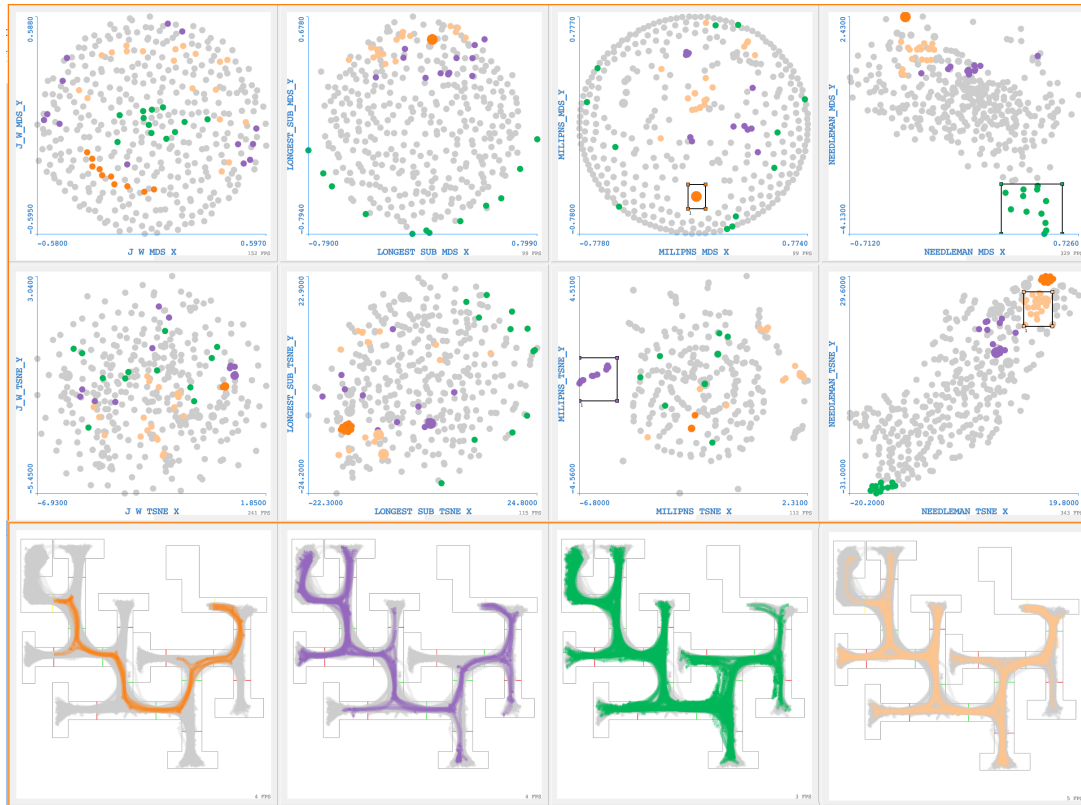
Figure 2

Figure 3: *MDS (top row) and t-SNE (middle row) embeddings using four proposed metrics are shown. We brush various clusters in single views and check if they stay together in all views. The metrics yield very different results. The bottom row shows corresponding animal trajectories for the brushed data.*

The purple brush shows a cluster in MLIPNS t-SNE embedding. Again, many views do not show the purple points together. It seems that the Needleman-Wunsch metric groups similarly long strings together. We will further explore the Needleman-Wunsch metric and use it for automatic clustering of the trajectories.

5. Conclusion

The extension of an interactive visual analysis tool has proposed by Bechtold et al. [BSM18] by integrating mechanisms for easy pattern finding as proven to be useful. Comparing sequences and finding paths which contain a specific sequence of traversed Multiple T-Maze gate segments is simple in the Sequence View. It allows more in-depth brushing of the data than current state of the

art methods. Though the search function is very rudimentary at the current state and can be improved to allow multiple and more complex search queries, it allows more deeper data exploration.

For the clustering of the Multiple T-Maze data we used four different metrics which compute the distance of the sequence, represented as a string, to the ideal sequence: the path from the start to the end area with no wrong turns. The currently used metrics result in only small clusters, which indicates that we have not found a suitable metric as of yet. But Needleman-Wunsch metric needs further investigation. Alternatively, clustering could be improved if it is not based only on the sequence string but some other criteria such as the distance between the paths or the divergence from the perfect, shortest path through the maze.

References

[AA13] ANDRIENKO N., ANDRIENKO G.: Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization 12* (01 2013), 3–24. doi:10.1177/1473871612457601. 2

[AAB∗13] ANDRIENKO G. L., ANDRIENKO N. V., BAK P., KEIM D. A., WROBEL S.: *Visual Analytics of Movement*. Springer, 2013. 2

[AAFG18] ANDRIENKO G., ANDRIENKO N., FUCHS G., GARCIA J. M. C.: Clustering trajectories by relevant parts for air traffic analysis. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan 2018), 34–44. doi:10.1109/TVCG.2017.2744322. 2

[Bir13] BIRREN J.: *Handbook of the Psychology of Aging*. Handbooks of aging. Elsevier Science, 2013. URL: https://books.google.at/books?id=KPlFBQAAQBAJ. 1

[BLHKN81] BUBNA-LITTITZ H., HOFECKER G., KMENT A., NIEDERMULLER H.: [Gerontological pilot study on learning ability and memory in the stressed rat (author's transl)]. *Aktuelle Gerontol 11*, 1 (Jan 1981), 28–31. 2

[BLJ94] BUBNA-LITTITZ H., JAHN J.: Psychometric testing in rats during normal ageing. procedures and results. *J Neural Transm Suppl 44* (1994), 97–109. 1

[BN06] BIZON J. L., NICOLLE M. M.: 32 - rat models of age-related cognitive decline. In *Handbook of Models for Human Aging*, Conn P. M., (Ed.). Academic Press, Burlington, 2006, pp. 379 – 391. URL: https://www.sciencedirect.com/science/article/pii/B9780123693914500333, doi:https://doi.org/10.1016/B978-012369391-4/50033-3. 1

[BSM18] BECHTOLD F., SPLECHTNA R., MATKOVIC K.: Visual Exploratory Analysis for Multiple T-Maze Studies. In *Eurographics Workshop on Visual Computing for Biology and Medicine* (2018), Puig Puig A., Schultz T., Vilanova A., Hotz I., Kozlikova B., VÃązquez P.-P., (Eds.), The Eurographics Association. doi:10.2312/vcbm.20181246. 1, 2, 3, 4

[CC00] COX T. F., COX M.: *Multidimensional Scaling, Second Edition*, 2 ed. Chapman and Hall/CRC, 2000. URL: http://www.amazon.com/Multidimensional-Scaling-Second-Trevor-Cox/dp/1584880945. 2, 3

[HCC∗19] HE J., CHEN H., CHEN Y., TANG X., ZOU Y.: Diverse visualization techniques and methods of moving-object-trajectory data: A review. *ISPRS International Journal of Geo-Information 8* (01 2019), 63. doi:10.3390/ijgi8020063. 2

[HEB∗00] HOEGER H., ENGELMANN M., BERNERT G., SEIDL R., BUBNA-LITTITZ H., MOSGOELLER W., LUBEC B., LUBEC G.: Long term neurological and behavioral effects of graded perinatal asphyxia in the rat. *Life Sciences 66*, 10 (2000), 947 – 962. URL: http://www.sciencedirect.com/science/article/pii/S0024320599006785, doi:https://doi.org/10.1016/S0024-3205(99)00678-5. 2

[HES∗06] HOEGER H., ENGIDAWORK E., STOLZLECHNER D., BUBNA-LITTITZ H., LUBEC B.: Long-term effect of moderate and profound hypothermia on morphology, neurological, cognitive and behavioural functions in a rat model of perinatal asphyxia. *Amino Acids 31*, 4 (Nov 2006), 385–396. URL: http://dx.doi.org/10.1007/s00726-006-0393-z, doi:10.1007/s00726-006-0393-z. 1

[KGMQ08] KIM J.-S., GRAČANIN D., MATKOVIĆ K., QUEK F.: Finger walking in place (fwip): A traveling technique in virtual environments. In *Smart Graphics* (Berlin, Heidelberg, 2008), Butz A., Fisher B., Krüger A., Olivier P., Christie M., (Eds.), Springer Berlin Heidelberg, pp. 58–69. 2

[KKS∗11] KOHLHAMMER J., KEIM D. A., SANTUCCI G., ANDRIENKO G., POHL M.: Solving problems with visual analytics. In *The European Future Technologies Conference and Exhibition 2011* (2011), Procedia Computer Science. 2

[LWF11] LUINE V. N., WALLACE M. E., FRANKFURT M.: Age-related deficits in spatial memory and hippocampal spines in virgin, female fischer 344 rats. *Curr Gerontol Geriatr Res 2011* (2011), 316386. URL: http://dx.doi.org/10.1155/2011/316386, doi:10.1155/2011/316386. 1

[MHRW04] MONTELLO D. R., HEGARTY M., RICHARDSON A. E., WALLER D.:. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2004, ch. Spatial Memory of Real Environments, Virtual Environments, and Maps., pp. 251–285. 2

[MWSB12] MATKOVIC K., WINDING C., SPLECHTNA R., BALKA M.: Interactive Visual Analysis of Ethological Studies: Getting Insight from Large Ensembles of Animals' Paths. In *EuroVA 2012: International Workshop on Visual Analytics* (2012), pp. 85–89. 2

[PJCK03] PARNPIANSIL P., JUTAPAKDEEGUL N., CHENTANEZ T., KOTCHABHAKDI N.: Exercise during pregnancy increases hippocampal brain-derived neurotrophic factor mrna expression and spatial learning in neonatal rat pup. *Neuroscience Letters 352*, 1 (2003), 45 – 48. URL: http://www.sciencedirect.com/science/article/pii/S0304394003010085, doi:https://doi.org/10.1016/j.neulet.2003.08.023. 2

[Rob07] ROBERTS J. C.: State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In *Proc. of the 5th International Conference on Coordinated & Multiple Views in Exploratory Visualization* (2007), IEEE CS Press. URL: http://www.cs.kent.ac.uk/pubs/2007/2559. 2

[SJL∗18] STEIN M., JANETZKO H., LAMPRECHT A., BREITKREUTZ T., ZIMMERMANN P., GOLDLÃIJCKE B., SCHRECK T., ANDRIENKO G., GROSSNIKLAUS M., KEIM D. A.: Bring it to the pitch: Combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics 24*, 1 (Jan 2018), 13–22. doi:10.1109/TVCG.2017.2745181. 2

[Sma01] SMALL W. S.: Experimental study of the mental processes of the rat. ii. *The American Journal of Psychology 12*, 2 (1901), 206–239. URL: http://www.jstor.org/stable/1412534. 1

[SN27] STONE C. P., NYSWANDER D. B.: The reliability of rat learning scores from the multiple-t maze as determined by four different methods. *The Pedagogical Seminary and Journal of Genetic Psychology 34*, 4 (1927), 497–524. URL: https://doi.org/10.1080/08856559.1927.10532397, arXiv:https://doi.org/10.1080/08856559.1927.10532397, doi:10.1080/08856559.1927.10532397. 1

[Str] TextDistance âĂŞ python library for comparing distance between two or more sequences by many algorithms. https://pypi.org/project/textdistance/. Accessed: 2019-03-07. 3

[SvL16] SLINGSBY A., VAN LOON E.: Exploratory visual analysis for animal movement ecology. *Comput. Graph. Forum 35*, 3 (June 2016), 471–480. URL: https://doi.org/10.1111/cgf.12923, doi:10.1111/cgf.12923. 2

[SW75] SIEGEL A. W., WHITE S. H.: The development of spatial representations of large-scale environments. vol. 10 of *Advances in Child Development and Behavior*. JAI, 1975, pp. 9 – 55. URL: http://www.sciencedirect.com/science/article/pii/S0065240708600075, doi:https://doi.org/10.1016/S0065-2407(08)60007-5. 2

[TC05] THOMAS J. J., COOK K. A.: *Illuminating the path: The research and development agenda for visual analytics*. IEEE, 2005. 2

[vdMH08] VAN DER MAATEN L., HINTON G.: Visualizing high-dimensional data using t-sne. 2, 3

[WPF09] WALF A. A., PARIS J. J., FRYE C. A.: Chronic estradiol replacement to aged female rats reduces anxiety-like and depression-like behavior and enhances cognitive performance. *Psychoneuroendocrinology 34*, 6 (Jul 2009), 909–916. URL: http://dx.doi.org/10.1016/j.psyneuen.2009.01.004, doi:10.1016/j.psyneuen.2009.01.004. 1