**TimeSpin: A tool for visualization of large spatio-temporal datasets**

by

Alexander Hohl

*Introduction*

The visualization of geographical data over time has become a hot topic in Geovisualization, especially if there are large (massive) volumes of data involved. However, our ability to collect and store spatio-temporal data has outpaced our ability to process and visualize it in recent years (Compieta et al., 2007). Challenges involve the design of static and animated visualizations that reduce cluttering and occlusion of data elements (Shepherd, 2008), excessive complexity of animation and the resulting cognitive load for viewers (Harrower and Fabrikant, 2008).

Space-time cube is a way to display spatio-temporal data that foregoes animation by the use of two spatial and one temporal axes (x, y, t), which allows the viewer to follow the trajectory of data objects through time (Nakaya and Yano, 2010). If space-time cube is used for displaying point data, it is essentially a simple 3D scatterplot, which is very challenging to the viewer and his/her ability to recognize spatio-temporal patterns due to cluttering and overprinting. Therefore, Delmelle et al. (2014) use space-time kernel density and isosurfaces in their visualization of the 2010 Dengue fever epidemic in Cali, Columbia. That way, spatio-temporal clusters of disease occurrence can easily be detected in a journal style static visualization.

*Related Work*

Shepherd (2008) compares a 2D view of family classes in London to 3D stacked symbol visualization. He argues that in two dimensional maps, with increasing amount of data, point symbols obscure each other (fig. 10.1 a). While this can be dealt with using transparency and cut-outs, the case when objects represent identical locations is much harder to solve. He uses the third dimension to stack symbols that share a location (fig. 10.1.b). This could be applied to the Dengue fever dataset, as the number of point symbols is high (9’500) and some share locations as they might be disease cases in the same household or block and because point locations were moved to the next street intersection to guarantee a level of geomasking. However, I don’t see the benefit of Shepherd’s symbol stacks approach, as the degree of symbol obscurity is not lowered in my eyes.

In order to track geographical changes over time, map animations can be powerful tools for visualization, if they are well designed and pitfalls are avoided (Harrower and Fabrikant, 2008). As remembering each frame of the animation puts a cognitive load on the viewer, an animation should not be too long, typically less than one minute. The passage of time is best visualized in a ‘temporal legend’, and the authors show three examples of it: digital clock, cycle and bar. Cycle and bar have the advantage of displaying current time while simultaneously showing the relation to other instants in time. Potential pitfalls include ownership changes of parcels, which are conceptualized as instants in time. A linear temporal animation is ineffective in that case as there are long spans without any events on the map, punctuated with periods of instantaneous change. A static map would be appropriate in that case.

In order to facilitate analysis of massive datasets, efficient and effective techniques for mining and analyzing spatio-temporal datasets are needed (Compieta et al. 2007). Spatio-temporal data mining uses a set of exploratory, computational and interactive approaches for analyzing these very large datasets. Visualization is recognized as to be powerful in this domain, since humans are able to visually detect and interpret patterns. Interactive representations are needed in the data mining process in order to obtain different spatial and temporal views of the data. Therefore, the authors present their Google Earth-based visualization tool, which allows users to interact with the visual representation of data and dynamically modify parameters to see how they affect the visualized data. The view panel shows google earth and the data associated with it, the data panel allows querying the data and therefore filtering what is being displayed and what not. The dimensional panel allows the user to move in four dimensions, namely 3 spatial ones and time. The ability to interact with the data representation is very important to me, especially when map animation is involved. Filtering is a nice way to temporarily reduce the amount of data displayed to ameliorate the problem of clutter/overprinting.

Tominski and Schuhmann (2008) present a spiral view for detecting cyclic patterns in datasets. The application is very interactive, as spiral width, center offset and cycle length are adjustable and doing so, patterns will eventually appear in the spiral. Subsets of the spiral can be displayed separately. The Dengue fever dataset could be visualized that way, as seasonality is always interesting in disease research. In order to satisfy my needs, there has to be a link to the spatial view of things, which could be done by an additional window.

Piringer, Kosara and Hauser (2004) investigate extensions to traditional 3D scatterplots to solve problems of perception and interaction while presenting their very own application. In order to improve depth perception, they use variable point size with smaller points being farther away from the user. In order to help differentiation between points, they use halos of lower brightness but same hue as the corresponding point. An interesting approach for displaying point density in the 3D scatterplot is presented: The authors draw 2D Histograms on each border plane (alternatively: heat maps), to show the point-distribution on the plane. This makes me think that they could have showed point density within the scatterplot, using the methods shown in Delmelle et al., 2014. Anyway, the user can choose a time interval (temporal focus), which then is subject to interaction while all the data that is not in the focus (context) is not and is displayed in grey color. In addition, the 3D scatterplot view is supported with three 2D scatterplots of orthographic projections of the data onto the border planes. This is also supported by the ability to define 3D brushes in a 2D environment.

*Data and Approach*

Here, I present an approach and tool to visualize spatio-temporal point data that reduces the symbol occlusion problem. In a first step, I plotted the Dengue fever disease cases (Delmelle et al., 2014) in a space-time cube as points (Fig. 1). The coordinates range from Xmin = 1’054’199 to Xmax = 1’068’485, from Ymin = 857’581 to Ymax = 878’930 and from Tmin = 3 to Tmax = 365 (Julian day of the year). There are three visual properties of the point symbols that can be altered: color, size and transparency. If these three properties are carefully combined, the visualization should be equal the one shown in Delmelle et al. (2014) in terms of the viewer’s ability to detect spatio-temporal patterns. The visual properties relate to the density values that are generated for each point through space-time kernel density estimation, using the same parameters as Delmelle et al. (2014). That way, high density points are red, big in size and have low transparency. Low density points are blue, small in size and have high transparency.

As symbols still occluded each other, I used a data mining approach to reduce the number of points to be displayed. I plotted disease frequency over time (Fig. 2) and visually identified phases of change in disease occurrence:

1. Emergence, 2) Climax, 3) Decrease

The idea is to focus on these 3 phases and display them distinctively in the same visualization. These time windows are most interesting to visualize as they most accurately represent the underlying processes that led to the spatio-temporal distribution of the Dengue fever cases. Since the choice of boundaries between the phases was made subjectively, I allowed them to be variable in the tool (b1, b2, b3), so that a potential user could specify them.

Finally, I defined a rotation-axis around which I rotated the points (Y = 880’000, also variable; Fig. 3). The rotation angle α of each point depends on its timestamp: the later the occurrence, the wider α (up to 360°). In order to keep the three phases separate from each other, I split the resulting rotation circle into three equal parts (0 – 120°, 121 – 240°, 241 - 359°) and performed the rotation of each phase in its sector. As shown in the *Results* section, the epidemic struck in the first third of the year, so I assigned an angle of 359° to all cases that happened later than that. That way, they are all on one plane, easily recognizable but don’t take much space in the plot or occlude other symbols. In order to maintain distinctiveness between the phases (as they are adjacent to each other), I used an exponential relationship between time and angle and scaled the angles so that their associated points remain within the sector. The exponential factor *p* can be adjusted and has an effect on distinctiveness of the three phases in the rotation plot (Fig. 4, Fig. 5) as it shapes the relationship between t and α.

This approach conserves the spatial dimensions of the area while the temporal axis is curved to a circle. That way, symbol occlusion is reduced and spatio-temporal relationships can be detected. I used MayaVi library for python scripting within the Enthought Canopy software environment for visualizations (<http://code.enthought.com/projects/mayavi/>).

*Results*

The space-time cube shows the distribution of Dengue fever disease cases of 2010 in Cali, Columbia (Fig. 1). Due to technical problems, I could not adjust the transparency of the symbols individually. Anyway, this static visualization is unsatisfactory since it is hard to make out spatio-temporal patterns and trends. We can see some high-density areas in the foreground, but we cannot relate them to each other spatially.

The three phases of the epidemic (b1, b2, b3, Fig. 2) have the following temporal ranges: 1) 3 - 45, 2) 46 - 65, 3) 66 - 130. The number of disease cases is distributed over the phases the following way: 1) 1584, 2) 1266, 3) 2800, and 3905 after phase 3. We basically see a strong increase of disease cases in the first phase, which slows down and turns into a decrease in the second phase. The third phase shows a slowly decreasing disease frequency. Only the first third of the year has to be displayed in order to show the distribution of Dengue fever cases during the 2010 epidemic. We also see a slight peak later in the year around T = 230.

The rotation plots (Fig. 4, Fig. 5), show the final distribution of the points. Using an exponential factor of *p* = 0.5 yields perfectly distinguishable phases (“nuclear hazard sign” pattern), a factor of *p* = 1 does not (“wheel” pattern). Those disease cases that happened later than the three phases can be seen as plains near 12 o’clock. We also see a gap between the second and third phases in both plots which might be caused by the narrowness of the chosen phase-boundaries. It’s a relative short period of time and even though this is the peak phase, it does not contain as many disease cases as the other two phases.

*Conclusion & Future Work*

TimeSpin needs more work, in order to contribute substantially to the set goal: to make it easier to detect spatio-temporal patterns in large geographical datasets. My personal impression is that after applying these geometrical conversions to the dataset, it has not become easier to read.

Therefore, future work includes optimizing the visual properties of the points. That way, they should be easier to discern and their location relative to each other should be clearer. In addition, the plot should contain a reference map for improved visibility. It could be on a semitransparent plane that is interactively rotatable. That way, the distribution of Dengue disease cases at any time can be displayed.

An additional analysis using different data would be helpful. Applying TimeSpin to other datasets ultimately tests its applicability in a general sense.

*Figures*

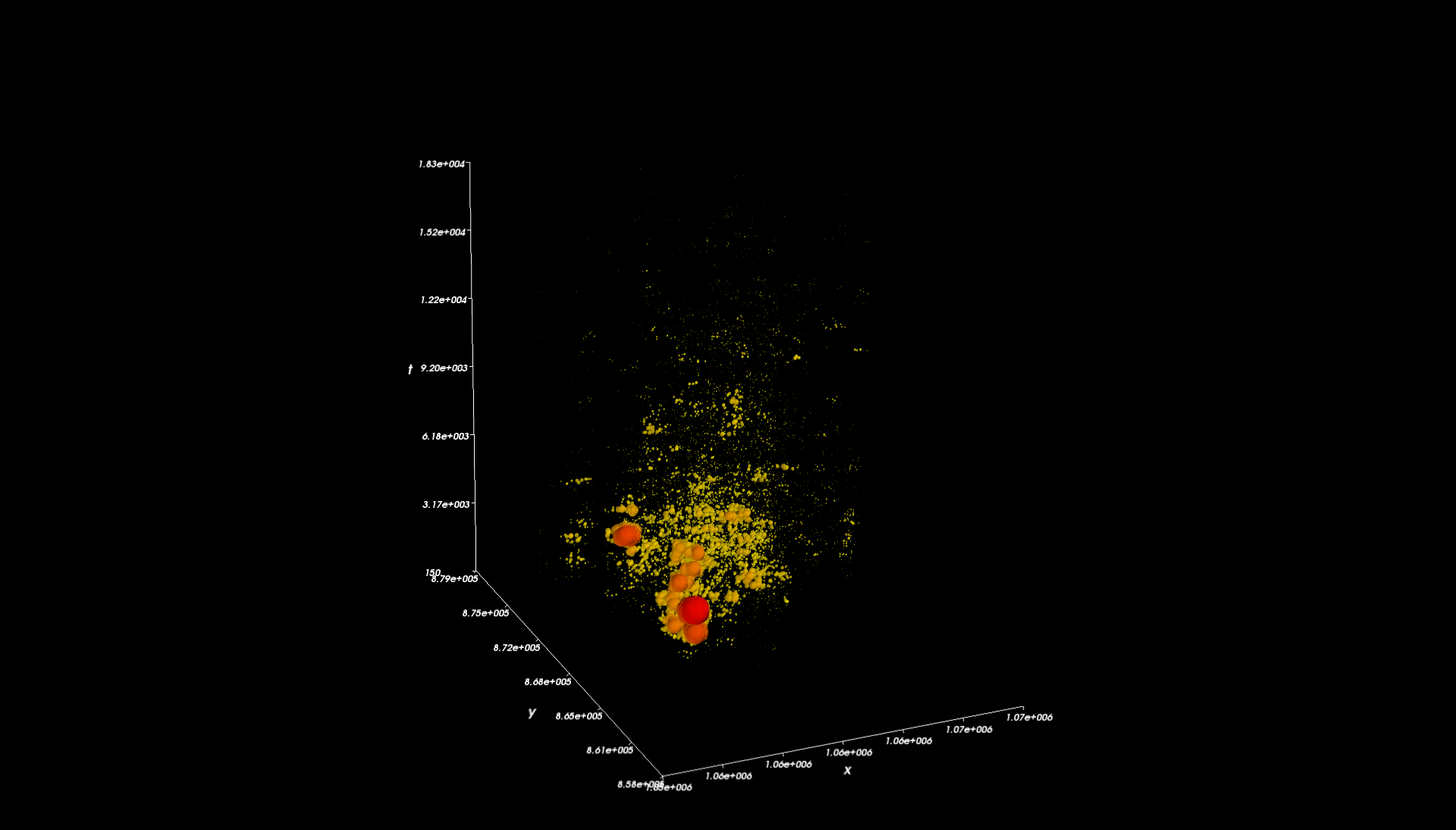
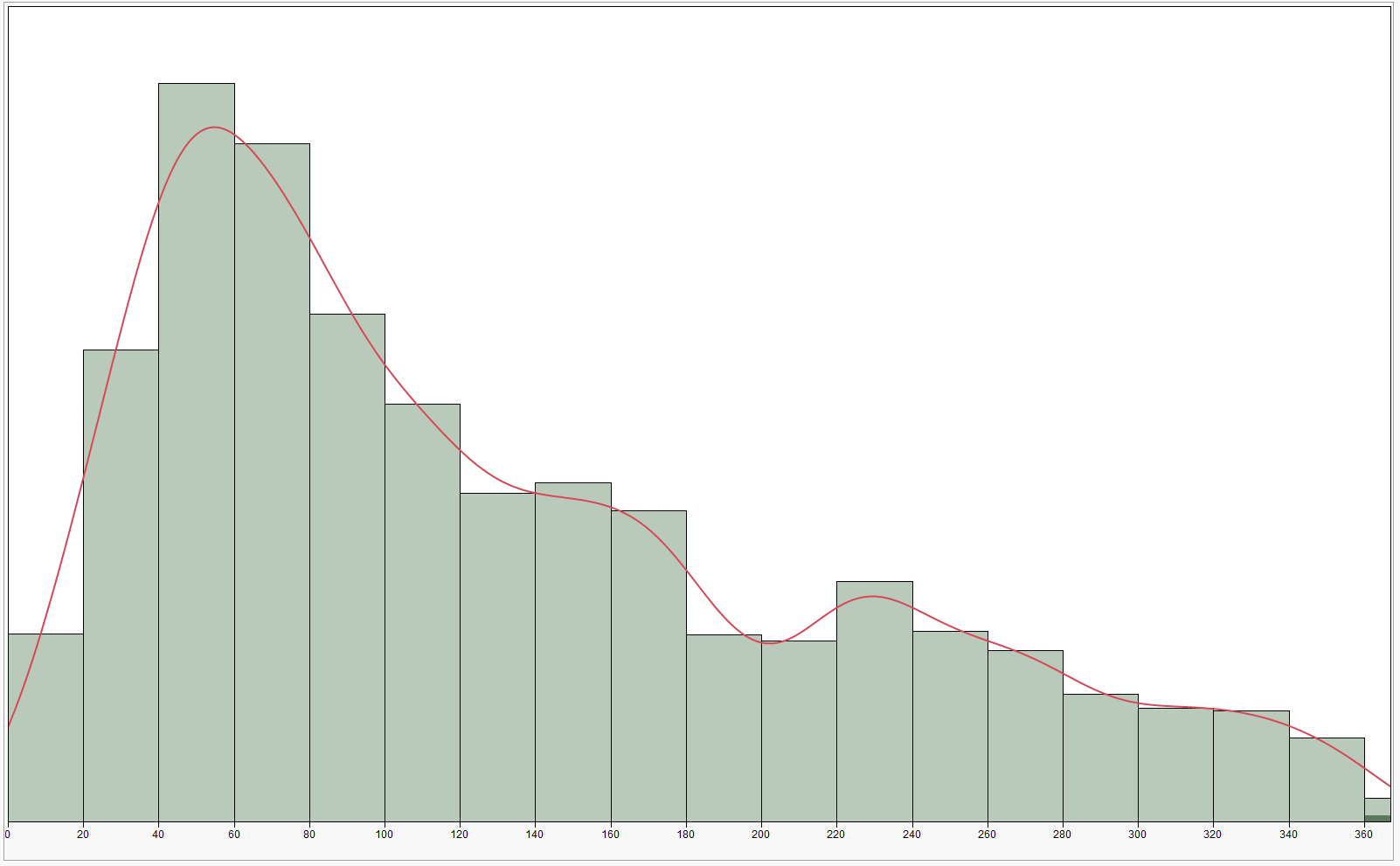
**

Fig. 1: Space-time cube of Dengue fever cases of 2010 in Cali, Columbia



**1**

**3**

**2**

Fig. 2: Three phases of disease in 2010: 1) Emergence, 2) Climax, 3) Decrease.

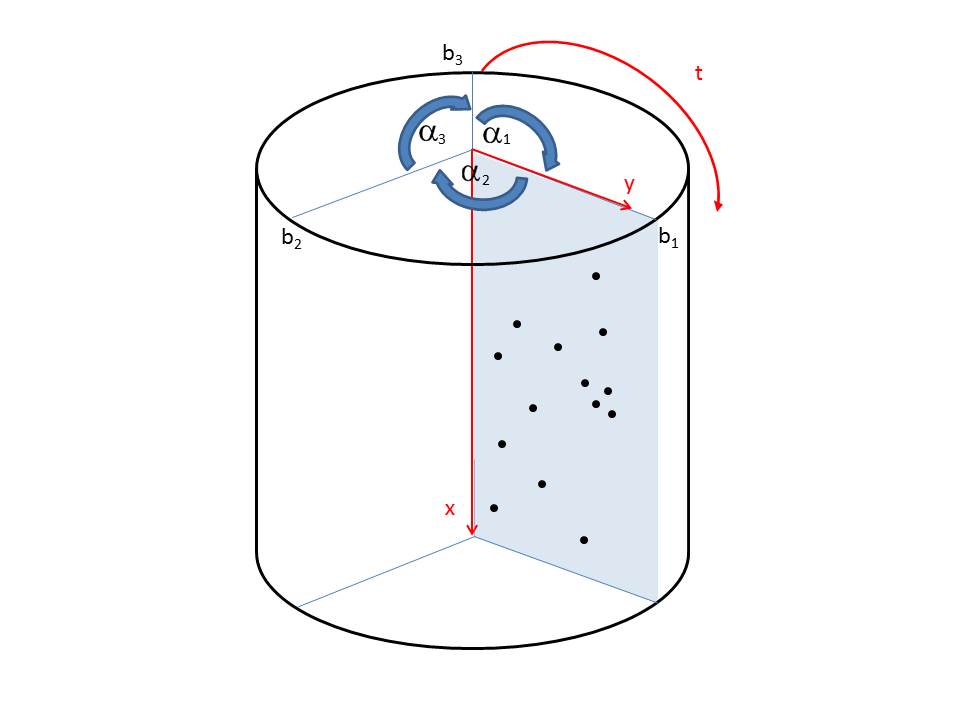
**

Fig. 3: Rotation, Angles, Sectors

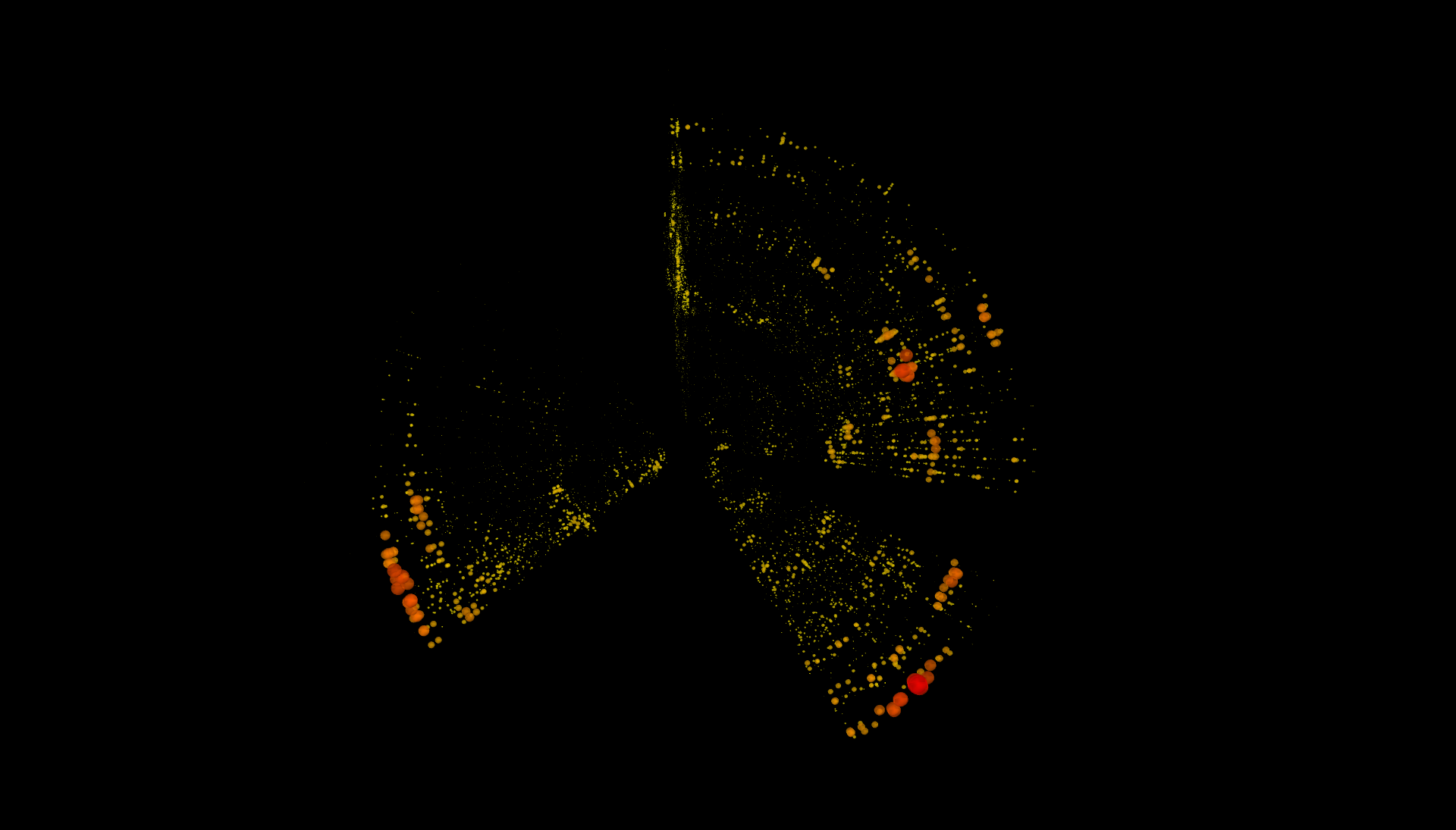


Fig. 4: Disease phases in rotation plot with *p* = 0.5.

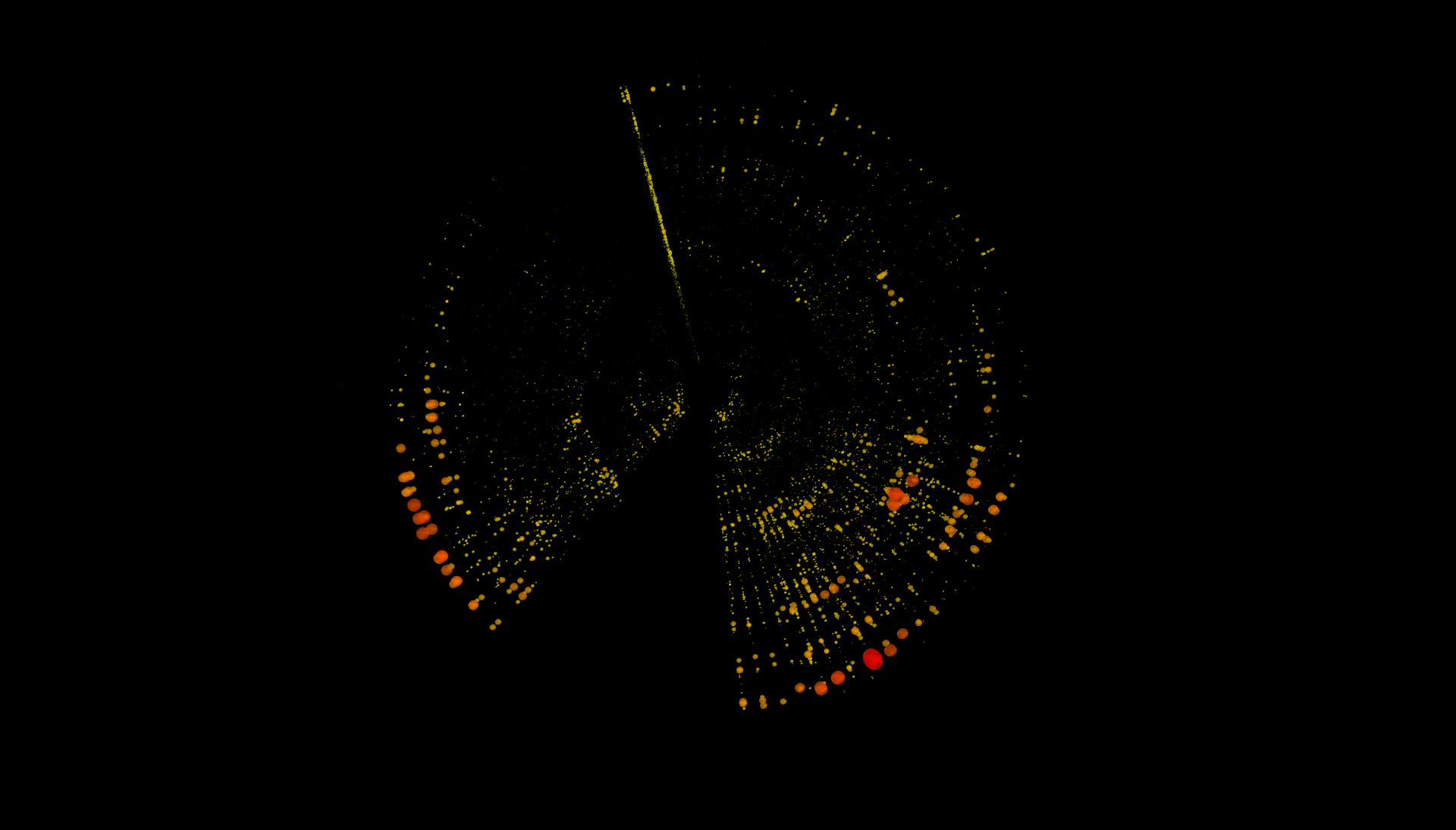


Fig. 5: Disease phases in rotation plot with *p* = 1.

*References*

Delmelle, E., Dony, C., Casas, I., Jia, M., & Tang, W. (2014). Visualizing the impact of space-time uncertainties on dengue fever patterns. International Journal of Geographical Information Science, 00(00), 1–21. doi:10.1080/13658816.2013.871285

Shepherd, Ifan D. H. (2008) Travails in the third dimension: a critical evaluation of three-dimensional geographical visualization. In: Geographic visualization: concepts, tools and applications. Dodge, Martin and McDerby, Mary and Turner, Martin, eds. John Wiley & Sons, Chichester, pp. 199-222. ISBN 9780470515112

Harrower, M. and S. Fabrikant (2008). The Role of Map Animation for Geographic Visualization. Geographic Visualization, John Wiley & Sons, Ltd: 49-65.

Compieta, P., Di Martino, S., Bertolotto, M., Ferrucci, F., & Kechadi, T. (2007). Exploratory spatio-temporal data mining and visualization. Journal of Visual Languages & Computing, 18(3), 255-279.

Tominski, C. and Schuhmann, H., 2008. Enhanced interactive spiral display. In: Proceedings of the annual SIGAD conference – special theme: interactivity. Sweden: Stockholm.

Nakaya, T. and K. Yano (2010). "Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics." Transactions in GIS 14(3): 223-239.

S. Walton and M. W. Jones, Interacting with Volume Data: Deformations using Forward Projection, Proc. 4th International Conference on Medical Information Visualisation -- BioMedical Visualisation, 48-53, IEEE July 2007