Mapping Contagion
Cartographic Visualizations of Time-Series Data

Andrew Tang
Information Design + Visualization
Northeastern University
MAPPING CONTAGION
Cartographic Visualizations of Time-Series Data

Thesis Presented by Andrew Tang
to

The College of Arts, Media, and Design
In Partial Fulfillment of The Requirements for The Degree of
Master of Fine Arts in Information Design + Visualization

Chair | Thomas Starr
Advisor | Dietmar Offenhuber
Readers | Lakshmi Goyal and Janos Stone
Respondent | Aleszu Bajak
Member | Nathan Felde

Northeastern University
Boston, Massachusetts
May, 2017
Abstract

Showing changes through time that are related to geographic space usually entails multiple views, either statically as small multiples or dynamically as an animated interactive. This thesis will attempt to present time-series events in a singular view, to show the breadth of the changes through time in one static view. This thesis will explore a space-time cube display and also the use of glyphs to convey the changes within an event. Being able to show time-series events statically and efficiently can greatly improve the spread of information, as the viewer can quickly grasp the entirety of the information being conveyed.
Acknowledgments

Christine Chan and Mom | giving me all the support in the world.
Omar, Tofu, and Danger | sometimes you just need to hug a cat.
Meredith Adinolfi and Keith Wollman | allowing me the flexibility to do what I do.
Andrew C. Smith | because he was there at the beginning and the end.
Pedro Cruz | for your exacting design & humor throughout our process.
A median network map showing location and frequency
The sizes of circles represent the frequency. Variants in black are from the United States, blue are Spain, and white are Japan (Martinez-Hernandez, F., et al., 2013, Virology Journal 7, p. 196).
Table of Contents

1 Introduction 3
2 Why Visualize 7
  2.1 Cartographic Space and Time 7
  2.2 Temporal Mapping 9
  2.3 Events, Slices, and Episodes 14
  2.4 Duration and Frequency 15
3 Contemporary Applications 17
  3.1 Influenza 18
  3.2 Rate of Incidence 24
  3.3 Spatial Maps 28
  3.4 Spatial Temporal Maps 35
4 Conceptual Framework 41
  4.1 Influenza 42
  4.2 Zika: Three Dimensions 49
  4.3 Influenza: Two Dimensions 56
  4.4 Zika: Two Dimensions 58
5 Conclusion 61
6 References 63
How long will ‘it’ take to get here? People want to know when it’ll get to where they are.

—Angela McLean

FIGURE 1
Qualitative outbreak reconstruction based on effective distance
A circular tree diagram mapping distance and time of an outbreak of an infectious agent through simulations of air travel where each node is an airport. (D. Brockmann and D. Helbing, 2013, Science 342 (6164), p. 1337-42.)

QUOTE
Coming to an Airport Near You
Introduction

Winter is coming. The brisk air seems colder as you rush off for your morning commute, but you’ve forgotten your scarf today. By midday you might feel a slight itch in your throat; your afternoon coffee is accompanied by a faint cough, but you try to suppress the oncoming threat by reassuring yourself that it’s passing because you have a big project due within the next week. You commute home with a growing headache and before you know it, you arrive home in concert with body aches, a sore throat, and fatigue. These are all symptoms of the common flu, Influenza. Influenza is a fast-acting virus that annually infects between five and twenty percent of the population. It is the most successful contagion in our history. We have probably all experienced it sometime in our lives as the common flu, but most people do not realize how deadly of a virus Influenza is. Approximately 200,000 Americans are hospitalized every year, and 3,000–49,000 die of flu-related causes (CDC, 2016).

Influenza is one of many infectious viruses that we as humankind have been exposed to, but with modern technology and learned experience throughout our history, we have been able to control the virus through greater knowledge of transmission and the invention of vaccines. Smallpox, another deadly virus that as recently as 1967 infected over 15,000,000, resulting in 2,000,000 deaths that year, has virtually been eradicated (WHO, 2016).

In the past, many viral infections have been localized to specific areas, such as the Cholera and Typhus outbreaks in London and other industrially dense cities throughout most of the 19th century but with an increasingly connected world the threat of contagion is becoming ever more present. The Influenza pandemic of 1918 is an extreme example of the accelerated diffusion of a virus into the population due to connectivity; 50,000,000
INTRODUCTION

to 100,000,000 were killed worldwide. As with Smallpox, vaccinations have sharply reduced the annual deaths of Influenza from what the world saw during the 1918 pandemic to an average of 23,000 deaths annually for the past 31 flu seasons. Even within this statistical average of 23,000, this doesn’t tell the complete story, as the fluctuations range from 3,000 in 1986 to 48,000 in 2003.

World War I greatly influenced the infection rate, with the massive amounts of troop movement made possible by technological advances and the introduction of modern-age travel. Visualizing simulations of air travel helps scientists understand the modern travel network to see patterns of transmission and directionality of the viral vector. Having additional context and gained knowledge of travel patterns assists in predictability of the movement. This knowledge can further inform protocols for controlling future outbreaks (Fig. 1).

In recent years, there have been new outbreaks and epidemics. In the 2014 Ebola outbreak in Western Africa, the remote geographic location of the outbreak limited the diffusion of the virus. Western Africa isn’t known to be an international hub for travelers and the knowledge base of the scientific community was able to identify the virus, prevent the spread, and maintain control of the spread. Technology and the connectedness of the global community played a positive role in this case. More recently, Zika was introduced to the world stage while ushering in the Rio 2016 Summer Olympics. With the world traveling to the games, diffusion of infected persons would be immense, but there was no catastrophic global outbreak after the Olympics. The Zika virus is still an extensive dilemma to solve, but global initiatives have already implemented response protocols.

Humans learned from past experiences and have a better understanding of how viruses are transmitted from person to person and how they move around the world within the travel network. Learning from past experiences is core to human nature; even molecularly, this is the main function of our immune system and is essential to the success of vaccines. Vaccines are pseudo-copies of viruses that train our immune system to recognize the pathogen and create antibodies that attach to the actual pathogens as targets for T-helper cells to dispose of. An increased understanding of viruses has led to a decrease in infectious diseases. The development of antibiotics and vaccines has plummeted the global rate of mortality from 0.8% in 1900 to less than 0.1% currently (Fig. 2).

The ability to visualize spatial-temporal data increases our knowledge of historical context, which gives a reference point to make comparisons to past experiences and gain insight and understanding into how the world works.

FIGURE 2
Rate of Infectious disease through history
Line graph showing deaths in the United States from various infectious diseases and initiatives that have taken place to combat the epidemics. (J.H. Strauss, E.G. Strauss, 2008, Viruses and Human Disease).
INTRODUCTION

Deaths per 100,000 People per Year

- 40 states have instituted health departments
- Influenza pandemic
- Last human-to-human transmission of the Plague
- First use of penicillin
- First continuous municipal use of chlorine in water in the United States
- Salk Vaccine introduced
- Start of the HIV/AIDS epidemic

Year

1900 1920 1940 1960 1980 2000
FIGURE 3
The lower Mississippi River valley
Mapping the meandering fluctuations of the Mississippi River using layers and colors to represent the changes over time. (Fisk, H.N., 1944, Geological Investigation of the Alluvial Valley of the Lower Mississippi River, U.S. Army Corps of Engineers).
Why Visualize

Visualizing data is important to quickly and clearly grasp the data at hand. It is a tool to help see the inherent qualities of the presented information, and it helps us see patterns and gains an understanding of large datasets. Visualizations should introduce information to the viewer that might not be apparent and therefore should start a dialogue for a deeper dive by increasing the viewer’s curiosity into the subject.

Information has always been stored and presented in tabular form, as spreadsheets are quite familiar to anyone who has ever been tasked with keeping records of any sort. The mapping of contagio, with incidence of infection and location as the two main variables of information to be catalogued, can easily take this form as a two-dimensional dataset. But data in tabular form can gravely misrepresent the inherent qualities of the data. Anscombe’s Quartet is an example of why visualization is an important step in quickly seeing these qualities. This dataset (Fig. 4) was created in 1973 by statistician Francis Anscombe to support the importance of graphing or charting tabular data. The dataset contains four separate scatter plots consisting of x and y coordinates. At first glance by looking at common summary statistics such as mean, median, or regression coefficient, the coordinates of all four plots in tabular form are indistinguishable from each other. But upon visualizing the data, it is visually apparent that each scatter plot has drastically contrasting qualities.

Cartographic Space and Time

One of the first examples of spatial mapping is Dr. John Snow’s map of London’s Cholera outbreak. Previous to his infamous 1854 publication, John Snow published his findings in 1849. His initial publication didn’t receive the recognition as his second due to the lack of visual presentation: his datasets were originally presented in tabular form. The datasets were
**Anscombe’s Quartet**

Francis Anscombe, a statistician formed this dataset in 1973 to demonstrate the importance of graphing tabular data. The results of graphing the dataset revealed patterns and trends in the data that would not be apparent otherwise.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>8.0</td>
</tr>
<tr>
<td>y</td>
<td>8.0</td>
<td>9.1</td>
<td>7.5</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>8.0</td>
<td>6.8</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>13.0</td>
<td>13.0</td>
<td>13.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
<td>8.7</td>
<td>12.7</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>9.0</td>
<td>9.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>8.8</td>
<td>8.8</td>
<td>7.1</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>11.0</td>
<td>11.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>8.3</td>
<td>9.3</td>
<td>7.8</td>
<td>8.5</td>
</tr>
<tr>
<td></td>
<td>14.0</td>
<td>14.0</td>
<td>14.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>10.0</td>
<td>8.1</td>
<td>8.8</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>6.0</td>
<td>6.1</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>6.1</td>
<td>6.1</td>
<td>6.1</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>4.3</td>
<td>4.0</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>3.1</td>
<td>5.4</td>
<td>8.8</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>10.8</td>
<td>9.1</td>
<td>8.2</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>7.0</td>
<td>6.4</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>4.8</td>
<td>7.3</td>
<td>7.0</td>
<td>6.9</td>
</tr>
</tbody>
</table>

**FIGURE 4**

Anscombe’s Quartet
constructed with columns consisting of Districts, Population, Deaths from Cholera, and Water Supply (Fig. 5).

In his second publication, John Snow mapped the thirteen wells in and around the Soho District, where the outbreak had its largest impact. Using the spatial depiction of the data he was able to create a visual comparison of the proximity of deaths to each of the wells. The visualization showed a greater number of deaths in the immediate area of the Broad St. pump, and the visual correlation made a compelling argument that Cholera is a water-borne disease (Fig. 6). Dismantling the handle of the pump helped to thwart the outbreak.

In 1964, the modern evolution of spatial mapping known as GIS (Geographic Information System) was conceived by Roger Tomlinson, CGIS (Canada Geographic Information System). CGIS was put into place to take inventory and to analyze usage of Canada's large land mass. The significant addition to the traditional map is that Tomlinson introduced a layered approach to differentiate geographic features. These features can include forest, rivers, and roads, but this schema also allowed layers to include additional information such as soil drainage, national parks, climate conditions, and even political boundaries. GIS was soon adopted by the United States Census Bureau in the 1970s.

Temporal Mapping
In 1944 Harold Fisk was tasked by the U.S. Army Corps of Engineers to investigate the sequence of events in the evolution of the lower Mississippi River in the Alluvial Valley. Prior engineering projects were deployed to provide water supplies to parts of the valley. Aerial photographs of the area taken made the study possible as the transformations of the river were translated to static visualizations to clearly depict the alternations propagated to the river and valley over the 15-year span (Fig. 3).

Temporal mapping has always been relevant in the GIS community. It is not a new phenomenon, but with the evolution of mapping technologies and the universal use of GPS (Global Positioning System), the variable of time is playing a large role in presenting location-based data because the instance of time is relevant for comparisons to other instance; without these comparisons, there is no context. Depicting temporal data as another variable in a visualization adds complexity. Incidence and location are the primary variables to visualize contagion, but to show diffusion of a viral vector, time must be accounted for to explore the sequence of proliferation. I will further refer to incidence in this paper as an attribute of the data consisting of quantity or intensity of infection. Location will reference specific x, y, coordinates of a geographic point of a cartogram, and time will consist of time-series data.

Depicting temporal data in cartographic space can further be defined by specific characteristics of the type of presentation that is used (Kraak, 1994). M.J. Kraak (PhD in Cartography, Delft Technical University) includes both static and dynamic depictions of temporal data into three categories.
### JOHN SNOW | CHOLERA: Table V

<table>
<thead>
<tr>
<th>District</th>
<th>Population in 1853</th>
<th>Deaths /100k</th>
<th>Water Supply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bermondsey</td>
<td>48,128</td>
<td>150</td>
<td>Southwark &amp; Vauxhall</td>
</tr>
<tr>
<td>S. Savior</td>
<td>35,731</td>
<td>146</td>
<td>Southward &amp; Vauxhall, Lambeth</td>
</tr>
<tr>
<td>S. George</td>
<td>51,824</td>
<td>143</td>
<td>Southward &amp; Vauxhall, Lambeth</td>
</tr>
<tr>
<td>St. Olave</td>
<td>19,375</td>
<td>134</td>
<td>Southward &amp; Vauxhall</td>
</tr>
<tr>
<td>Rotherhithe</td>
<td>17,805</td>
<td>112</td>
<td>Southward &amp; Vauxhall, Kent</td>
</tr>
<tr>
<td>Whitechapel</td>
<td>79,579</td>
<td>95</td>
<td>East London</td>
</tr>
<tr>
<td>Newington</td>
<td>64,816</td>
<td>57</td>
<td>Southwar &amp; Vauxhall, Lambeth</td>
</tr>
<tr>
<td>Kensington</td>
<td>73,699</td>
<td>53</td>
<td>WestMiddlesex, Chelsea, Grand Junction</td>
</tr>
<tr>
<td>Wandsworth</td>
<td>50,764</td>
<td>51</td>
<td>Southwark &amp; Vauxhall, Pumpwells, River Wandle</td>
</tr>
<tr>
<td>St. George</td>
<td>48,376</td>
<td>43</td>
<td>East London</td>
</tr>
<tr>
<td>Camberwell</td>
<td>54,667</td>
<td>40</td>
<td>Southward &amp; Vauxhall, Lambeth</td>
</tr>
<tr>
<td>Stepney</td>
<td>110,775</td>
<td>34</td>
<td>East London</td>
</tr>
<tr>
<td>Lambeth</td>
<td>139,325</td>
<td>34</td>
<td>Lambeth, Southwark &amp; Vauxhall</td>
</tr>
<tr>
<td>Greenwich</td>
<td>99,365</td>
<td>31</td>
<td>Kent</td>
</tr>
<tr>
<td>Marylebone</td>
<td>157,696</td>
<td>30</td>
<td>West Middlesex</td>
</tr>
<tr>
<td>Westmer</td>
<td>65,609</td>
<td>27</td>
<td>Chelsea</td>
</tr>
<tr>
<td>St. James</td>
<td>36,406</td>
<td>25</td>
<td>Grand Junction, New River</td>
</tr>
<tr>
<td>Hackney</td>
<td>58,429</td>
<td>22</td>
<td>New River, East London</td>
</tr>
<tr>
<td>Paddington</td>
<td>46,305</td>
<td>22</td>
<td>Grand Junction</td>
</tr>
<tr>
<td>Shoreditch</td>
<td>109,257</td>
<td>21</td>
<td>New River, East London</td>
</tr>
<tr>
<td>Bethnal Green</td>
<td>90,193</td>
<td>20</td>
<td>East London</td>
</tr>
<tr>
<td>Poplar</td>
<td>47,162</td>
<td>17</td>
<td>East London</td>
</tr>
<tr>
<td>West London</td>
<td>28,840</td>
<td>14</td>
<td>New River</td>
</tr>
<tr>
<td>Hanover Square</td>
<td>33,196</td>
<td>12</td>
<td>Grand Junction, New River</td>
</tr>
<tr>
<td>and May Fair</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Islington</td>
<td>95,329</td>
<td>12</td>
<td>New River</td>
</tr>
<tr>
<td>Chelsea</td>
<td>56,538</td>
<td>11</td>
<td>Chelsea</td>
</tr>
<tr>
<td>East London</td>
<td>44,406</td>
<td>9</td>
<td>New River</td>
</tr>
<tr>
<td>London City</td>
<td>55,932</td>
<td>9</td>
<td>New River</td>
</tr>
<tr>
<td>Clerkenwell</td>
<td>64,778</td>
<td>8</td>
<td>New River</td>
</tr>
<tr>
<td>Belgrave</td>
<td>40,034</td>
<td>7</td>
<td>Chelsea</td>
</tr>
<tr>
<td>St. Martin-in-the-Fields</td>
<td>24,640</td>
<td>5</td>
<td>New River</td>
</tr>
<tr>
<td>St. Pancras</td>
<td>166,956</td>
<td>5</td>
<td>New River, Hampstead, West Middlesex</td>
</tr>
<tr>
<td>St. Luke</td>
<td>54,055</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 5**

*John Snow’s 1849 tabular dataset of the Cholera outbreak (J. Snow, 1849, *Medical Gazette and Times)*

**FIGURE 6**

*John Snow’s 1854 spatial mapping of the Cholera dataset (J. Snow, 1854, *Medical Gazette and Times)*
Change is the essence of cartographic time

—Gail Langran
FIGURE 7
Graphic depiction of categorization of time
Geographic space is represented on the x and y axis, and time is represented on the z axis. (G. Langranm 1989, Time in Geographic Information Systems).

QUOTE
Time in Geographic Information Systems
**SINGLE STATIC MAP**
A single static map represents a single event or episode. Graphic variables are used to reproduce information displaying value, size, and directionality.

**STRIP MAP**
The previously coined term “strip map” can be grouped into the visualization technique called small multiples. Small multiples are datasets that are partitioned by category or grouping. Each filtered dataset is then graphed or charted separately but is shown in a collective of the whole dataset to be easily compared between the single graphs, charts, or, in this case, maps. Each individual map in a small multiple is essentially a single static map since the graphic elements are used to show the same variables while time is portrayed in the sequential succession in the collective of the small multiples.

**ANIMATED MAP**
An animated map is a dynamic version of a strip map, but at each distinct view it acts like and presents the same information as the single static map. The visual variables are the same as the single static map and is not used to portray time. Temporal information is depicted through the animation.

**Events, Slices, and Episodes**
The increased need to represent time as a variable has led the GIS community to further abstract and define the effects of time in cartographic space. Gail Langran (PhD, University of Washington) extends cartographic theory past the traditional visual language of using graphic symbols to represent changes over time. By defining relative domains to create a shared language between spatial and temporal data, he has introduced dynamic relationships by categorizing cartographic time as events.

A change of any significance can be categorized as an event. Events as a unit of time can be represented as slices or episodes (Fig. 7). Slices are the fundamental unit of an event that catalogs a specific instance, while episodes are series of slices delineated in chronological order. To further define this logic and to classify dynamic time, Peuquet introduces World state and Change state. A World state is the spatial distribution of a given phenomenon at a given time, a slice. A Change state accounts for change during a given time span, an episode, but additionally defines classes within the Change state as: continuous, majorative, sporadic, and unique (Peuquet, D.J., 2005).

- Continuous—going on throughout some interval
- Majorative—going on most of the time
- Sporadic—occurring some of the time
- Unique—occurring only once
Defining temporal change adds values to the specific characteristic of each classification and significantly gives relevance to duration and frequency.

**Duration and Frequency**

Duration and frequency are inherent qualities in measuring and monitoring contagion. Duration consists of the life span and an epidemic, and there are many hidden factors that compose this. The timeline of this life span does not just start at the tipping point of the diffusion of an infectious virus, but at the onset of an initial infection. The following steps are a guideline for the detection and control of an outbreak, while each step can vary depending on resources, but a main factor for the duration of an outbreak mainly depends on the control of the outbreak itself.¹

- Establish the existence of an outbreak
- Define what constitutes a case and identify cases as they occur
- Formulate hypotheses on the causes, and implement initial control measures
- Prepare for an emergency
- Test the hypotheses through analysis of surveillance data or special studies
- Draw conclusions and re-adjust hypotheses and control measures if needed
- Prevent spread of the disease in the community
- Maintain essential services
- Implement, test, and revise the national plan for long-term prevention and control

The CDC (Center for Disease Control) defines frequency as a measure of central location that provides a single value that summarizes an entire distribution of data. Common frequency measures are ratios, proportions, and rates.²

**Ratio:** The relative magnitude of two quantities or a comparison of any two values. It is calculated by dividing one interval- or ratio-scale variable by the other. The numerator and denominator need not be related. Therefore, one can compare any types of quantities.

**Proportion:** The comparison of a part to the whole. It is the ratio in which the numerator is divided by the denominator. You might use a proportion to describe what fraction of clinic patients tested positive for a vector or what percentage of the population is younger than a certain age.

**Rates:** A measure of the frequency with which an event occurs in a defined population over a specified period of time. Rates put disease frequency in the perspective of the size of the population. Rates are particularly useful for comparing disease frequency in different locations, at different times, or among different groups of persons with potentially different sized populations.

²CDC 2012, “Principles of Epidemiology in Public Health Practice.”
The geography of talk in Great Britain
This figure shows the strongest 80% of links, as measured by total talk time, between areas within Britain. The opacity of each link is proportional to the total call time between two areas, and the different colors represent regions identified using network modularity optimization analysis (Ratti, C., et al., 2010, PloS One, p. e14248).
Contemporary Applications

The criteria that I will use to analyze the following contemporary applications will consider the use of visual language within each of the visualizations, and how the graphic variables are used to depict the rate of incidence of infection, the use of cartographic time, and the use of cartographic space to ease the understanding of the information and if all aspects—rate, time, and space—are present in the visualization. The following applications all contain rate of incidence, so the differences will be how the design has portrayed the visual elements and visual metaphors of representing cartographic space and time. What methods of visual storytelling were used to their advantage and disadvantage?

Each of the next sections will increase the complexity of the previous section by adding a variable that is needed to show changes of events through time. The first section will give examples of how rate of incidence has been portrayed. The next section will add cartographic space as a variable to be depicted as a graphic element. In the third section, cartographic time will be the last variable introduced, to create visualizations that contain all three aspects of a location-based, time-series event.

Rates are usually defined with lines, shapes, and color. Differences in representation of cartographic space mainly depends on what the final message can afford—whether the communication needs to be static and printed or dynamic or interactive as an animation. Representing cartographic time is a larger design challenge than the other two variables, as it depends on both of the previous. If a graphic primitive such as color is used to depict rate, then can hues of that color be used to show changes in time. Or will that confuse the viewer, or will the change be too subtle to recognize? Also, the final deliverable of the visualization will dictate whether time can be represented by using small multiples or animation.
Rate of Incidence

This visualization is a tree map but, more specifically, a phylogenic tree. The inherent quality of a phylogenic tree affords visualizing the divergence of growth. New lines show an emergence of a new vector, while the length of the line can indicate a variety of information, i.e., distance, quantity, or time. This is a traditional representation of phylogenic tree that is used in science illustration. Here the horizontal axis depicts an episode of time ranging from 1944 to 2008. All three phylogenic trees show a continuous state of change. Genetic variance and diffusion is shown along with location, but cartographic space is not used. The branches show emerging variations in the genome of the Dengue virus.

Advantages
- able to show a large amount of information in a small space
- able to easily compare length of lines for comparison of time
- able to quickly get a sense of what the data is represent

Disadvantages
- no representation of geographic space

**FIGURE 8**
The phylogenic relationships of DENV-1 genotypes
A traditional phylogenetic tree used in science, showing sequences of Dengue virus genotypes. (Villabona-Arenas et al., 2013, PloS One, p. e62649).
Here to is a phylogenic tree. But in this case, color has been introduced as a value to distinguish groupings and to categorize. Time is still represented by the length of the line, but is not displayed horizontally, making the comparison of the length of gene lines very difficult. Additionally, the x,y position of the lines can show locality or directionality, but those variables are not present in this visualization.

Advantages
- use of color to denote groupings of information
- visually appealing

Disadvantages
- no direct representation of geographic space
- directionality of lines have no inherent meaning
- hard to compare the line lengths

**CONTEMPORARY APPLICATIONS**

**FIGURE 9**

Evolutionary history of DENV-1
This visualization is also a tree map, but a circular tree map where the length of the line depicts distance whereas the two previous applications are using line length to portray time. The purpose of this visualization is a simulation to show the speed of diffusion of a viral vector. The origin or primary location is at the center of the circular tree map. In this specific simulation, the biological viral vectors are shown diffusing through air travel, so the locations depicted are airports. The node of color shows when that specific location becomes infected. The colors help make a comparison of which locations will become infected in a given time span.

**Advantages**
- able to show a large amount of information in a small space
- able to easily compare length of lines for comparison of distances
- added color helps make comparisons of diffusion
- able to quickly get a sense of what the data is representing
- visually appealing

**Disadvantages**
- no direct representation of geographic space

**FIGURE 10**
Qualitative outbreak reconstruction based on effective distance
At first glance this visualization can be mistaken for a circular tree graph. But a radial plot is used here to show the gene flow of the influenza virus (Fig. 11). The center location of the plot is indicated by the main location label; the spokes of the radial plot account for the other locations the information is being mapped against, and the concentric circles along the radial plot indicate the value of the gene flow. Color is used to differentiate either viral protein or enzyme flow.

**Advantages**
- able to show a large amount of information in a small space
- able to compare values along concentric circles
- use of icons to denote a different variable of the information
- visually interesting

**Disadvantages**
- no direct representation of geographic space
- the mental model of the visualization does not match the metaphor for representation, so it takes additional capacity to fully understand what the visualization is representing

**Figure 11**
*Gene flow of countries with higher number of influenza cases*
Radial plots used to show gene flow influenza human virus A H1N1. (Martinez-Hernandez et al., 2013, Virology Journal 7, p.196)
Information graphics, data graphics, data visualization, data art. Is there a difference, or is the difference only defined by the domain where the information is being presented (Lev Manovich 2011)? This popular visualization (Fig. 12, D. McCandless, 2016) is a simple line chart that uses the X, Y, and Z axes to show three variables. The X axis represents time, the Y axis represents quantity, and the Z axis along with color indicates a different category. Here, McCandless is making a statement of media coverage and actual deaths for each category. This visualization is an example of creating a visually effective narration of a topic.

Advantages
- able to see the increase and decrease of various sets of time-series information
- uses color to show multiple datasets
- able to quickly get a sense of what the data represents
- able to easily see trends and patterns
- visually appealing

Disadvantages
- difficult to make exact comparison of the data

---

**Figure 12**
Mountains Out of Molehills — Information Is Beautiful
Spatial Maps
When mapping the growth and spread of any global variable, cartographic space adds contextual information. Fig. 13 illustrates the potential spread of two biological vectors, H1N1 and SARS (D. Brockmann et al., 2013). In panels A and , the map projection is used to give global context. Panel A uses color as a value of different viral vectors and lines to show diffusion of the various strands, while panel B only uses color dots mapped to the projection to simulate initial global outbreak. Panels C, D, and E are scatterplots that show distance traveled over time. Panel C creates a comparison overview of seven different viruses, and panels D and E focus on H1N1 and SARS, adding a dimension of the quantity of cases with the size of the diameter of the dots.

Advantages
· presence of geographical space to show contextual location
· using color to show and compare different viral vectors
· able to quickly get a sense of what the data represents

Disadvantages
· needs multiple views/panels to show the data
· needs interactivity to fully explore the data
This cartographic representation shows multivariate data. The half circles depict levels of potential abundance of *Ae. aegypti*: none, low, moderate, and high. The top half contains data from the month of January, and the lower half is data collected from the month of July. The sizes of the circles represent the average number of arrivals into the United States by persons visiting from Zika-infected countries. Dotted lines and color show variables for area and the map projection displays locational information (Fig. 14).

**Advantages**
- able to see the increase of potential abundance from January to July
- presence of geographical space to show contextual location
- using color to show multiple variables for additional information
- able to quickly get a sense of what the data represents
- able to easily see trends and patterns
- visually interesting

**Disadvantages**
- difficult to make size comparisons

**FIGURE 14**
Arrival by air and land from countries on the CDC Zika travel advisory.
A multivariate cartogram showing location of the number of visitors from Zika-infected countries. (A. Monaghan, 2016, PLoS Currents 8 (March)).
Spatial Temporal Maps
This visualization is an example of strip maps, currently known as a small multiple. Geographic space anchors the visualization, while time is shown on sequential maps displaying a different episode of time. (Fig. 15).

Advantages
· able to see changes through time
· presence of geographical space to show contextual location
· use of color to show growth or spread of incidence
· able to quickly get a sense of what the data represents
· able to easily see trends and patterns

Disadvantages
· space needs to be considered for displaying many instances of small multiples

FIGURE 15
This visualization shows spatial temporal relationships by using color and hue as values. This map is split into area by county: color is used to show the different regions. Each region is further separated into slices to depict changes over time. Within each of those slices, hue is used to show the rate of capacity. More color is used to show the regional high school for each of the counties. This map was created to be a multivariate visualization (Fig. 16).

Advantages
- many variables can be represented
- presence of geographical space to show contextual location

Disadvantages
- hard to understand what is being represented
- multiple uses of color create confusion
- using hue to show changes of time adds another color scheme. which adds to the confusion
This visualization utilizes the space-time cube to depict the density of tanker movements in space and time. Cartographic location is depicted on the x,y axis, while time is conveyed on the z axis. Here, color is used as a value in coordination with opacity to show traces of movement (Fig. 17).

Advantages
- presence of geographical space to show contextual location
- cartographic time and space is represented

Disadvantages
- the use of opacity and color makes it difficult to see changes through time
- needs interactivity or animation to see totality of visualization
- three-dimensional visualizations incur occlusion of information

FIGURE 17
Time space cube
An important element of human representation of the world around us is the retention of information relating to past events.

—D.J. Peuquet
Conceptual Framework

Distilling information is a constant act in our current world. There is an abundance of information and data presented to us through various forms throughout our day. The influx of information has brought new ways to see the information, and influx in technology has increased the different avenues through which information can be presented. Processing all the information is key to understanding how we fit in with the world around us. This knowledge enables humans to effect the world in many ways, positively or negatively. So the value of being able to distill vast amounts of information to quickly understand and grasp the message is key to many avenues of communication.

Visual depictions of geographic information or maps are something that almost everyone is familiar with and can understand and grasp quickly. Maps show information about a geographic area that ranges from topographic information of terrain to road maps that can help you navigate from point A to point B, or even a map that contains a listing of amenities surrounding your current location. But maps can do more than merely serve as a way-finding device. Geographic information can be used to create public policy or used to communicate scientific research. Adding time-series data as contextual information can differentiate maps to have usage beyond way-finding.

My thesis focuses on static cartographic visualizations depicting episodes of temporal data by distilling the dynamic essence of events into a singular contextual view. Having a snapshot that contains temporal context to present patterns or trends of the current situation may give the viewer a better grasp of the information being shown. Contextual information will assist with the decision-making process, as it gives states of comparison to the current data. Contextual-temporal data affords
### CARTOGRAPHIC VARIABLES

<table>
<thead>
<tr>
<th>Overall configuration</th>
<th>Cartographic Space</th>
<th>Cartographic Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configurations separated by...</td>
<td>map</td>
<td>state</td>
</tr>
<tr>
<td>Regular sampling units...</td>
<td>sheet lines</td>
<td>events</td>
</tr>
<tr>
<td>Meaningful units</td>
<td>cells</td>
<td>decades, days, hours, etc.</td>
</tr>
<tr>
<td>Separators between unites</td>
<td>objects</td>
<td>versions</td>
</tr>
<tr>
<td>Size measured by...</td>
<td>boundaries</td>
<td>mutations</td>
</tr>
<tr>
<td>Position described by...</td>
<td>length, area</td>
<td>duration</td>
</tr>
<tr>
<td>Contiguous neighbors...</td>
<td>coordinates</td>
<td>date</td>
</tr>
<tr>
<td>Maximum number of neighbors</td>
<td>adjacent objects</td>
<td>previous and next versions</td>
</tr>
<tr>
<td></td>
<td>infinite</td>
<td>two</td>
</tr>
</tbody>
</table>

### GRAPHIC VARIABLES IN HIERARCHICAL ORDER OF PERCEPTUAL TASK

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>position</td>
<td>position</td>
<td>position</td>
</tr>
<tr>
<td>length</td>
<td>density</td>
<td>color hue</td>
</tr>
<tr>
<td>angle</td>
<td>color saturation</td>
<td>texture</td>
</tr>
<tr>
<td>slope</td>
<td>color hue</td>
<td>connection</td>
</tr>
<tr>
<td>area</td>
<td>texture</td>
<td>containment</td>
</tr>
<tr>
<td>volume</td>
<td>connection</td>
<td>density</td>
</tr>
<tr>
<td>density</td>
<td>containment</td>
<td>color saturation</td>
</tr>
<tr>
<td>color saturation</td>
<td>length</td>
<td>shape</td>
</tr>
<tr>
<td>color hue</td>
<td>angle</td>
<td>length</td>
</tr>
<tr>
<td>texture</td>
<td>slope</td>
<td>angle</td>
</tr>
<tr>
<td>connection</td>
<td>area</td>
<td>slope</td>
</tr>
<tr>
<td>containment</td>
<td>volume</td>
<td>area</td>
</tr>
<tr>
<td>shape</td>
<td>shape</td>
<td>volume</td>
</tr>
</tbody>
</table>

**FIGURE 18**
Cartographic Variables
Variables showing relational domains of cartographic space and cartographic time. (Langranm 1989, *Time in Geographic Information Systems*)

**FIGURE 19**
Graphic Variables
Variables defined by J.D. Mackinlay by hierarchy of perceptual task. (J.D. Mackinlay 1999, *Readings in Information Visualization: Using Vision to Think*)
meaningful context. A scatterplot with a single point has no message, provides no information, and has no importance because there are no other points of reference for comparison. “Similarly, events never occur in isolation, so when presenting singular slices of time there’s a loss in relational context from past data and new data,” (Krstajić, 2013). The ability to display temporal context gives the viewer a quick snapshot of the totality of an occurrence, an understanding of the whole. The value of this visualization benefits situations when timeliness is fundamental.

The static component of my thesis addresses the accessibility of a printed page. A printed page can be distributed easily and printed fairly cheaply anywhere. Animations and interactives need certain technologies to show the information. But the viewer needs to invest time either to explore and interact with the visualization or must spend time to watch the animation; either way, there is a threshold that the viewer must overcome.

Being able to present geographic time-series information in a compact manner is a valuable tool that can be used by many communities. The scientific community comes to mind when primary research needs to be communicated and ease of legibility and convenience of time are priorities. Relating research to the general public prioritizes ease of legibility. Complex ideas must be translated to match the mental model of the general public’s knowledge of the scientific content to make it accessible and therefore ingestible. Simplifying the data to see the inherent trends and patterns can create a more accessible document for the general public. The convenience of time affords communicating ideas at conferences full of posters, when your message must be communicated in a split second or risk being lost in the crowd. Again, simplifying your information to create accessible patterns and trends allow the viewer grasp the message quickly.

The cartographic variables defined by Gail Langran (Fig. 18) categorize the relationships between cartographic space and cartographic time. I will explore some of these categories to create my visualization. Cartographic space will be depicted by longitude and latitude coordinates but simplified into a spatial grid of regular units that resembles the shape of a specific area.

The variables for cartographic time that I will explore are: events, sampling units, duration, date, and previous and next versions. I will define events as a snapshot in the lifespan of a specific infectious disease. The sampling units from these events are units of time, which are dependent on collection and reporting intervals, such as weekly or monthly intervals. Duration applies to the time span of an episode within an event. And visualizing previous events add temporal context, which is the focus of my thesis.

The graphic variables I will explore have been detailed by Jock D. MacKinlay (Fig. 19) to explain the hierarchy of visual perceptions. I will focus on quantitative techniques, as infectious disease data
consist mainly of rates of infection, death rates, and prevalence, which all are quantitative information. Tamara Munzner’s table of expression of effectiveness (Fig. 20) gives visual examples of MacKinlay’s rankings of hierarchical perceptions.

Using position as a common scale gives a reference point for comparison of the data shown. Using length to show a quantity is similar to using a bar chart or graph to show distinct values and to easily make comparisons within the dataset.

Angle can be used to show the increase or decrease and the changes between each of the sampling intervals. The slope of an angle relates directly to the rate of increase or decrease. Area and volume will directly be tied to the rate of infection.

I will use color, hue and texture, for any additional information present in the dataset or used to create visual cues to help with the previous argument of making comparisons within each of the datasets.

**FIGURE 20**

**Effectiveness of visual expression**

Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels. (T. Munzner 2014, *Visualization Analysis & Design*).
The data I collected consisted of time-series information, which contains rate of infection or incidence rate of any virus or disease. Having location data is also very important, as I want to include geographic information in the visualization. So I found data that fit my criteria in two databases: the CDC (Centers for Disease Control) and GFT (Google Flu Trends). The resolution of the location data ranged from individual countries, to regional divisions, and state and city entities. I focused on the state level, as this give me a common baseline to create comparisons between a variety of datasets.

- Influenza | GFT (Google Flu Trends)
- Zika | CDC (Center for Disease Control)

*Influenza:* By collecting influenza data, I wanted to show patterns in something that is relatable to the general public. Influenza, a seasonal flu, is already well understood by the general public, so it’s not so foreign to make a connection.

*Zika:* I also wanted to visualize something that’s current and in the news, but something that’s not so familiar to the public. “Zika” seems like just a buzzword that is ignored since we hear it so much in the media. So comparing influenza, a common and familiar virus, to Zika not fully understood virus, would be a nice contrast to see what inherent qualities emerge from the visualizations.

What was missing from the data that I was hoping to find was any network data about directionality of diffusion. Most data that may contain information of diffusion is simulated data as the resolution of tracking the spread of individual viruses through carriers do not exist. You can only see if the rate of infection increases in a specific area, but you cannot assume where or how an area got infected because you can not assume coorelation. I used a variety of software to create the visualizations. Different types of software used to explore, while others, were used to create the final visualization style and others were used to translate the raw visualization into a presentable manner—basically the steps of parsing and organizing the data into digestible information. Here is the list of various softwares and applications.

- Tableau | explore data
- D3.js | explore data
- Google Sketchup | create visualizations
- Adobe Illustrator | create and polish the visualizations to create the visual collateral for presentation
Experiment 1

Event | Influenza
Sampling Unit | Weekly update by State
Duration | September 28, 2003 - August 09, 2015
Source | Google Flu Tracker

EXPLORING
In my initial experiment, I’m using Google Flu Tracker data. The dataset contains weekly incidence rates from 2003-2015 for the United States, and also includes granular information for geographical regions and specific notable cities, which resulted in a CSV file with 96,000+ cells. I filtered out all the granular data to focus on the 50 states and ended up with a dataset containing 31,500+ cells. I started to explore the data to see if there were any initial trends or patterns. The interesting trend you see reflects seasonal spikes in the rate of infections. You can also easily compare the seasonal spikes and pick out lower incidence seasons versus comparably higher incidence, for instance, the winter of 2013 looks like the highest rate of infection in recent history.

Advantages
· cartographic time is represented
· easily see patterns and trends
· use of line length to show duration
· use of angles to show frequency

Disadvantages
· no geographic information
· no cartographic space
FIGURE 21
Infection rates of influenza by state
Rate of infection is expressed on the y axis; time was shown with radius of the circles and color hue.
ORGANIZING

After exploring the data, I attempted to assign the variables of rate of incidence, time, and location, to different graphic variables. I placed rate on the y axis, and time was represented by radius of circle and also hue. I then projected each of the visualizations to the centroid of its state and onto a Robinson Projection (Fig. 21). This visualization was not at all successful; however, the drawbacks of this attempt did lead me to change the graphic variables on which I would map the data.

Advantages
- cartographic space is represented

Disadvantages
- difficult to understand what the data shows
- using hue to show time made it hard to distinguish the separate bins of time
- there is no common baseline to make comparisons of rate

ITERATION

After I made the change and assigned time to the y axis, and used the radius of the circles to represent rate of weekly infections (Fig. 22), the characteristics of duration and frequency became apparent.

FIGURE 22
Infection rates of influenza by state
Time is represented by the z axis, rate of infection is expressed by size of radius.
Experiment 2

Event | Zika
Sampling Unit | Weekly update by State
Duration | February 03, 2016 - March 08, 2017
Source | CDC

Continuing from my previous volumetric depiction, I used my Zika data to completely visualize the infection rates by travel of the United States. Instead of using the Robinson Projection to introduce geographic space, I created a grid to normalize the size and area of each state to help with the problem of occlusion when creating a three-dimensional graphic. Occlusion still occurs, but it’s less apparent in the northeast, where the physical areas of each state are smaller. In this visualization, you can see that each ring represents a change in the infection rate. If there are no changes in the number of infections, you are able to see a longer duration between the rings.

I wanted to try to create a static three-dimensional view that would convey the information truthfully because there is value in creating visual interest with your visualization. However, visualizing in three dimensions always has many drawbacks: occlusion of data between foreground and background, difficulty comparing volume, and the absence of a static baseline to make comparison.

Advantages
- cartographic time is represented
- cartographic space is represented
- quickly recognize patterns and trends

Disadvantages
- occlusion of data due to dimensionality
- no common baseline to easily make comparisons
After creating my space-time cube model, I knew I wanted to create a physical object for our thesis exhibition to explain my methodology of how I am talking about cartographic space and time.

*Laser:* I first explored laser cutting sheets of acrylic so each sheet would represent a slice of time. But after a few tests, I could not obtain the clarity to actually see the shapes of the visualization due to the refraction of light from the many surfaces of the sheets/slices, while trying to create an episode of time.

*3D printing:* I then went down the road of trying to create a mold to pour liquid acrylic to create a solid clear block, but the limiting factor was creating the shapes of the visualization with the 3D printing materials available. All the materials I had available would not sustain it’s shape due to the stress of pouring the liquid acrylic and the vacuum chamber needed to draw out the air bubbles to create the clarity I wanted.

*3D etching:* The solution I found was actually the easiest for me to create the exhibition piece because most of the work I needed to do was already done. I found a shop in Maryland that was willing to help me create my exactly what I had envisioned from the beginning. The method they used is three-dimensional glass etching. I first sent my 3D models to them to see if it was even compatible with thier equipment and after a quick test, I had the green light. I then adapted my model to the existing crystal sizes they had and started to think about an additional information that would be needed as a final piece.

I decided to keep the piece relatively clean due to the nature of the material. So I only added two lines to define the temporal context, the baseline date of February 16, 2016 to make comparisons of when the first infection was recorded with in the specific state, and the date at the end of my data collection, March 08, 2017.

The final piece arrived five days later and the detail and quality I was able to obtain with this technique was much better than expected. I would have never been able to represent my data as accurately with any of the other techniques I explored.
Experiment 3

Event | Influenza  
Sampling Unit | Weekly update by State  
Duration | September 28, 2003 - August 09, 2015  
Source | Google Flu Tracker

Due to the limitations of visualizing in three dimensions, I started to explore the same method of visualizing time on the y axis but in two-dimensional form. Within this two-dimensional view, I was able to start introducing color and hue to represent aspects of the data. I used hues of gray to create an even division of the time scale. You can quickly see comparable time frames, and compressing the time frames, you can see an outlier (red line), as the pattern of infection differed from all others. Adding the horizontal and vertical guidelines gives a reference point to make other comparisons within the data and to illustrate common infection rates throughout the full duration.

Advantages

- cartographic time is represented
- easily see patterns and trends
- use of hues to show episodes of duration
- able to compare frequency between episodes
I adapted the two-dimensional visualization to my Zika data and applied the same geographic area grid to apply cartographic space. Rate of infection is expressed on the x axis and time on the y. You can quickly spot the states with the most infections. You can see which states were infected slower than others and which had no infections i.e., Alaska. Another trend you can easily see is that there was an uptick in infection rates in August 2016 (red line).

The next steps to explore this visualization would be to intersect this dataset with other demographic data:

- normalize the data by population density
- juxtapose international airport locations and travel rates
- create direct comparisons to other viruses within the same timeframe

I’ve collected other dataset, including measles, and HIV, so I would like to apply the same visualization technique to these other datasets to see whether there are any intersecting or dissecting trends.
Changes through time influence how we understand the world, and previous experiences dictate how humans gain information. The retention of temporal context as an aspect of historical context affords humans a comparison of the current situation with previous experiences and can lead to a better understanding and decision-making process.

Using time-series data to visualize cartographic time and space must take into account all the variables involved to correctly depict the data: rate of incidence, time, and location. Many visualizations can successfully convey the rate of incidence with time or the rate of incidence with location. But creating a visualization that can depict all three variables is a challenge.

Expressing time and space has taken many forms: static maps, small multiples, and animations. Each of these forms has advantages and disadvantages in what types of information or data it’s visual language can communicate. Static maps can only convey a singular unit of time or a slice, while small multiples can show episodes of time. But area on the printed page can become an issue if many slices are needed. Animations take additional technology and time to understand and might not be accessible for mass consumption, so there is a threshold to overcome in any of these visualization types.

I have proposed a visualization technique that can express rate, time, and location. Considering the hierarchical order of perceptual task, I utilized the top quantitative graphic depictions—position, length, angle, area, and volume to express visual effectiveness. I used position to create a spatial grid for location, length to show incidence rate, and angle and slope to show degrees of increase or decrease. And I have only introduced color to create visual cues to highlight interesting qualities of data.

I believe my technique conveys the data concisely and effectively, allowing for any viewer to quickly grasp what the data is conveying by showing the patterns and trends for temporal context in a single static view.
References


