Roman Shaikh

Student number: 18300989

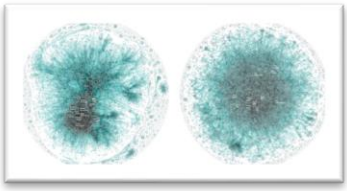
CS7DS4-Data Visualization

Assignment 3: Mid-term Written Assignment

Temperature and Mortality of London. William Farr, Report on the Mortality of Cholera in England, 1840-1850. London, 1852

Farr 1840 [https://goo.gl/eN2TjS]
Hi-Res [https://goo.gl/RnzzYs]

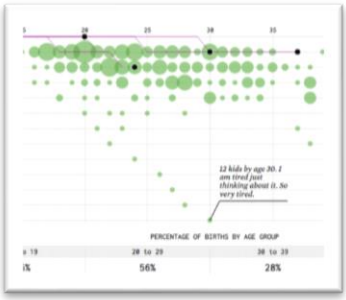
IDIOM	Circular Plot
Data Types	The main data types used for this diagram are • **Temperature:** Average/Mean Temperature of the week (in ° Fahrenheit), a Discrete numeric value. • **Mortality rate:** number of deaths registered from cholera, a Discrete numeric value. The full data is available as **Cholera** in **HistData** package in R. Data source: https://rdrr.io/rforge/HistData/man/Cholera.html
Tasks	**Compare:** • British epidemiologist William Farr's circular plots compares the cycles of temperature and cholera deaths in the 1840s. in London. **Correlate:** • The visualization tries to draw a casual correlation between the temperature and the mortality rate from cholera. **Identify:** • The chart by epidemiologist identifies the mortality rate during the period of 11 years from 1840-50 in London. Although the conclusion of Farr was wrong, this visualization has helped in establishing the strong importance of data visualization in the field.
Visual Encoding Channels	The circular plot uses 2 major encoding channels to show the relationship between temperature and the spread of the cholera epidemic. Color and position. **Position:** • The concentric circles define a scale for measuring two things 1 mortality and 2 the temperature. The distance between each circle represents either of two things 100 deaths or 10 degrees of temperature. • Also, there are two more circles one on the outside for average mortality(deaths) per year and one on the inside for average temperature that year. **Color:** • The color black on the outer circle shows the extent by which the weekly deaths exceeded the average mortality rate and the color yellow shows the amount by which the deaths per week are below the mean. • The Red color on the inside shows the level by which the average temperature of the week is more the yearly mean. • And the Blue color (not clearly visible as blue in the chart but specified in the source text) on the inside shows the amount by which the temperature that week was below the yearly temperature.



Comparing the last 10 years of patents filed at Apple and Google: Periscopic

Periscopic [https://goo.gl/HviByg]
Hi-Res [https://goo.gl/wfbuBe]

IDIOM	NODE-LINK DIAGRAM / Network stem diagram.
Data Types	The main data types used for this diagram are • **Patent:** A patent owned by the company conjunction with the inventor. Discrete value, uniquely identifiable by a patent number. • **Patent inventor:** The actual inventor of the patent product/idea. Discrete value uniquely identifiable by name • **Coinventors:** The co-inventors of the patent product/idea. Links: Coinventors link. Discrete value, uniquely identifiable by name. Links to other co-inventors. Exploring Innovation Signatures in Patent Data, Citraro 2017[https://goo.gl/itvYCL]
Tasks	**Cluster:** • Each dot/blob in the diagram represents an individual who is a patent inventor. • Dots that are clustered tightly indicate groups of inventors who work closely together. **Associate:** • And the links between each node represents a shared patent between the co-inventors. This associates the relationship between the inventors and coinventors (i.e. who controls the majority of the patents in an organization) **Correlate:** • Find the correlation between the patent owners and the organization structure. This may suggest the cultural differences between the two organizations.
Visual Encoding Channels	Various encoding channels have been used in this visualization to represent the relationships between data. **Size:** • The Size of the node/blob represents the number of patents the inventor has. **Position:** • The position of the nodes in each of the diagram shows the concentration of the majority of the patent holders in the organization. Dots that are clustered tightly indicate groups of inventors who work closely together. • This gives us an idea about the cultural difference in both of the companies. **Color:** • The blobs for founders and co-founders of the company are indicated in green to make them stand out and show. • The color intensity of each blob is in the proportion of the size giving us more idea about the number of patents the individual has.



How many kids do women give birth to and when do they have them?

Yau 2019. [https://goo.gl/zBokmX]
Source code [view-source:http://flowingdata.com/projects/2019/more-kids-timeline/]

IDIOM	Timeline
Data Types	The main data types used for this diagram are • **Number of Births:** The number of kids a woman gives birth to, a **discrete** numerical value. • **Age:** The age at which the women is at that particular childbirth, a **discrete** numerical value. • **Case ID:** uniquely identify each of the 1000 samples. **Derived data:** • **PERCENTAGE OF BIRTHS BY AGE GROUP:** The percentage of births grouped in age groups of 10-19, 20-29, 30-39 and 40-49 National Survey of Family Growth [https://goo.gl/PkwnJH] Data source [https://goo.gl/5Y4w5o]
Tasks	**Compare:** • The visualization compares the number of births each woman gives to their age. **Finding Distribution:** • The visualization shows the general distribution of birth-giving age of women is concentrated in the 20s. and as the number of kids grows the distribution shifts. **Categorize:** • Also, the visualization categorizes the women by age group i.e. 10-19, 20-29, 30-39 and 40-49 which contains the percentage of each women giving birth. **Highlight Outliers:** • The visualization also shows some outliers like the one woman with 12 kids.
Visual Encoding Channels	**Size:** • The green circles represent the total count for the given age. The bigger the circle a greater number of mothers gives birth at that age. **Position:** • The horizontal axis represents the age of the woman and the vertical axis represents the number of births. As we move along the horizontal axis we see the distribution of birth rate decreases i.e. less woman give birth after their twenties and also after the 2nd or 3rd child. **Motion:** • Each moving dot represents a mother, and with each birth the woman gives the dot moves down a unit (which is the number of child's)

Real-time Passenger Information (RTPI) for Dublin Bus

Source [https://goo.gl/wh6ZhF]

This data is provided by the National Transport Agency for tracking the real-time information of the Dublin Bus service.

I think it would be interesting to visualize this transport information in real-time using different encoding channels rather than traditional routes on a geographic map.

Data and data types: This dataset consists of numerous data which provide information on various data points. The most useful among these which will give most information to users would be as follows.

- **Origin**: The origin of the particular bus, A categorical value of the name of the origin stop e.g. College green.
- **Destination**: The destination of the bus, A categorical value of the name of the origin stop.
- **Due time**: The time the bus is due in at a particular stop, A Continuous value of time e.g. next 5 mins.
- **Arrival date time**: Time to destination for the bus, A Continuous value of time e.g. next 5 mins.
- **Scheduled arrival date time**: The Scheduled arrival time for the bus, A Continuous value of time e.g. next 5 mins.
- **Route:** Full route of the bus, A categorical value of the name of the origin stop e.g. Drumcondra
- **Direction:** the direction of the bus on the route, A Binary value for the bus direction E.g. UP or DOWN.

Tasks: Various visualization tasks can be achieved by visualizing this dataset. Some of the examples are as below.

- **Compare:** This visualization can help compare distances between different stops and the time it takes to reach it on an average.
- **Categories:** Can help classify different routes into categories like busy, not busy. Categorization can also help in better visualizing different bus service providers if considered.
- **Associate:** It can also help to draw a relationship between the time taken on average, traffic and the number of commuters on the routes.
- **Locate:** The visualization can help locate the current position of the bus on the route.

Encoding channels: The most effective channels for encoding information can be used to visualize this data.

- **Position:** Position can be used to show the current position of the bus relative to the bus stop.
- **Color:** Color can be used to categorise different routes in the city. Shades of can also be used to show the density of travel over a period of time.
- **Size:** The size of the indicator can be used to show the busyness of the route/ bus.
- **Motion:** Motion can show the current movement of the bus/traffic on the route.