Homework - 2

Name: Saumya Vora, Section:01, Course ID:IE 6600.18004.202010

1. Mapping Variables onto Aesthetics

Variables	**Type**	**Aesthetics**
Numeric	Continuous	Position, Size, Line width, Color
Numeric	Discrete	Position, Shape, Line Type
Categorical Unordered	Discrete	Shape, Line Type
Categorical Ordered	Discrete	Shape, Line Type
Date or Time	Continuous	Position, Size, Line Width, Color
Date or Time	Discrete	Shape, Line Type
Text	None or Discrete	Shape, Line Type

2. Explain the use of Sequential color scale and diverging color scale with examples.

Both the Color scales are used for representing Data Values.

Uses:

1) Sequential Color Scale:

- It is used to represent the data values such as income, temperature, speed.
- The visualization completely show which values are larger and smaller and how distant two specific values are from each other.
- For example, The Wind speed for geographic area can be visualized using sequential color.
- The ColorBrewer Blues scale is a monochromatic scale that varies from dark to light blue. The Heat and Viridis scales are multi-hue scales that vary from dark red to light yellow and from dark blue via green to light yellow, respectively.
- So, using monochromatic scale, dark blue for extreme high data value of wind speed and light blue for extreme low data value. We can represent using this scale in a map.

2) Diverging Color Scale:

- It is used to represent data values which deviate in one of the two directions relative to a neutral midpoint.
- To represent the dataset containing both positive and negative values diverging color scale is needed.
- For example, Temperature is measured in both positive and negative values. To represent "Temperature" geographically for United States.
- Common color choices for diverging scales include brown to greenish blue, pink to yellow-green, and blue to red. So, for negative extreme temperature could be identified by Blue color and positive extreme can be identified by Red color.

3. What type of variables can be encoded in a bar plot? What are the points to avoid while creating bar plots?
Types of Variables:

1) Categorical Variable
2) Numeric Variable
3) Date, Time
4) Text

- Points to avoid while creating bar plots

1) Labels should be avoided rotating when there is less space as it results awkward and difficult to read, instead just flip the axis.
2) Arranging bars alphabetically or arbitrarily should be avoided like descending order of the lengths of titles.
3) Sorting the plot by bar height makes no sense and hence should be avoided.
4) Grouped bar plot which shows lots of information can be confusing and hence should be avoided.

4. When does one use dot plots? What are the points to avoid while creating dot plots?

If the bars are too long and nearly same in the length, then eye is drawn to the middle of the bars rather than to the end points and hence fails to convey its message. So, for this kind of datasets it is impractical and may obscure key features. In these cases one uses dot plots.

Avoid

- Ordering the graph by alphabetical order(ascending or descending) for label names, as it causes dots to form a disordered cloud of points. This makes the figure difficult to read.

5. What are the pros and cons of pie chart, side by side bar plot, and stacked bar plot? Is a pie chart or stacked bar plot better than side by side bar plot for comparison?

Pie Chart:
Pros:

- Clearly visualize the data as a proportion of a whole.
- Visually emphasizes simple fractions, such as 1/2,1/3,1/4
- Look visually appealing even for very small datasets

Cons:

- The proportions must be add to 100%.
- Does not allow easy visual comparison of the relative proportions.
- Does not work well when the whole is broken into many pieces
- Does not work well for the visualization of many sets of proportions or time series of proportions

Stacked Bar Plot:
Pros:

- Clearly visualize the data as a proportion of a whole.
- work well for the visualization of many sets of proportions or time series of proportions

Cons:

- Does not visually emphasizes simple fractions, such as 1/2,1/3,1/4
- Does not Look visually appealing even for very small datasets
- Does not allow easy visual comparison of the relative proportions.
- Does not work well when the whole is broken into many pieces

Side by Side Bar Plot:

Pros:

- allow easy visual comparison of the relative proportions.
- Look visually appealing even for very small datasets
- Work well when the whole is broken into many pieces

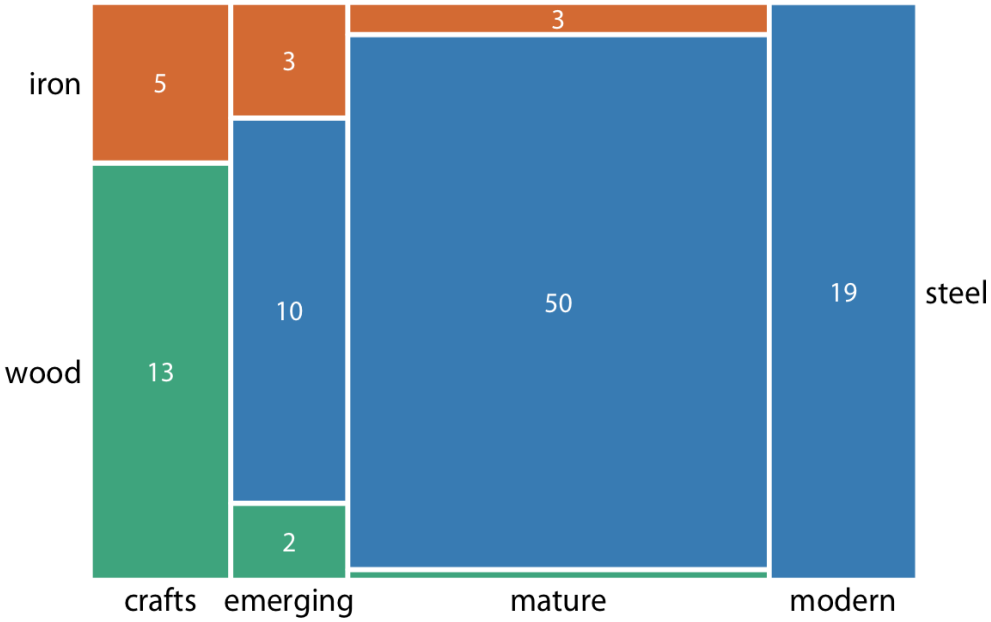
Cons:

- Does not clearly visualize the data as a proportion of a whole.
- Does not visually emphasizes simple fractions, such as 1/2,1/3,1/4
- Does not work well for the visualization of many sets of proportions or time series of proportions

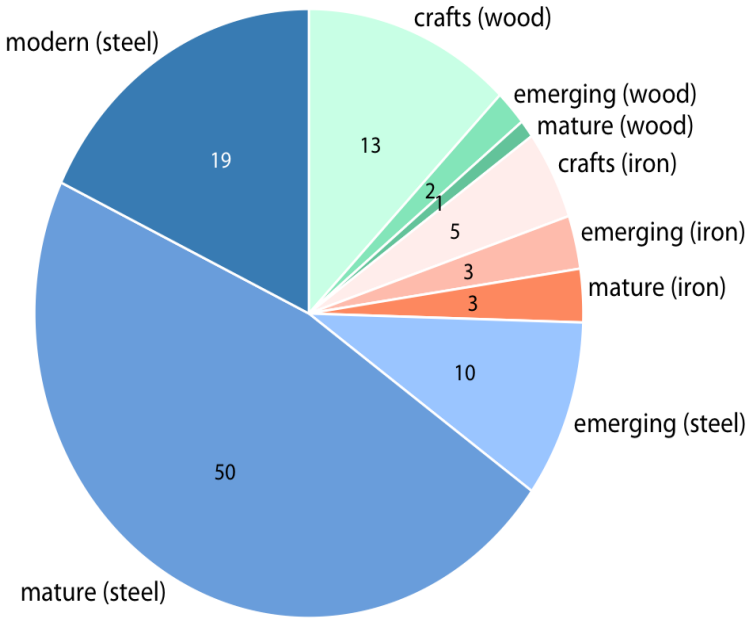
6. Give examples of visualizing nested proportions

Examples of visualizing nested proportions:

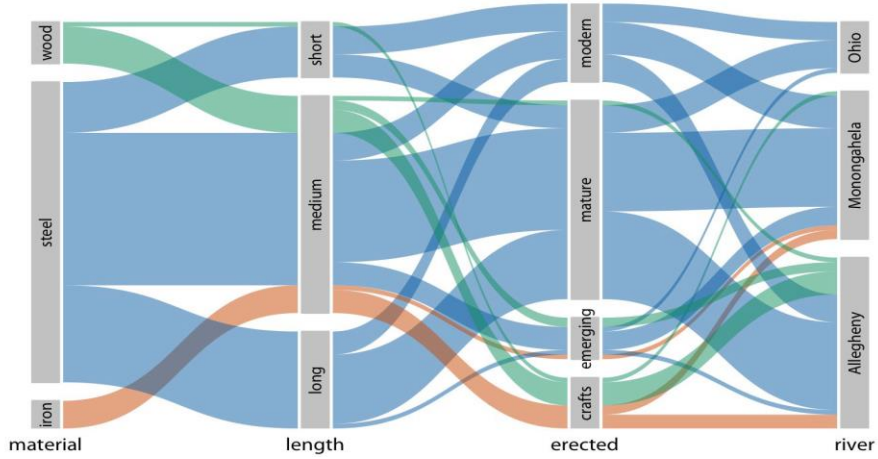
• Breakdown of bridges in Pittsburgh by construction material (steel, wood, iron) and by era of construction (crafts, emerging, mature, modern), shown as a mosaic plot. The widths of each rectangle are proportional to the number of bridges constructed in that era, and the heights are proportional to the number of bridges constructed from that material. Numbers represent the counts of bridges within each category. Data source: Yoram Reich and Steven J. Fenves, via the UCI Machine Learning Repository (Dua and Karra Taniskidou 2017).



• Breakdown of bridges in Pittsburgh by construction material (steel, wood, iron) and by era of construction (crafts, emerging, mature, modern). Numbers represent the counts of bridges within each category. Data source: Yoram Reich and Steven J. Fenves, via the UCI Machine Learning Repository (Dua and Karra Taniskidou 2017).



• Breakdown of bridges in Pittsburgh by construction material, length, era of construction, and the river they span, shown as a parallel sets plot. The coloring of the bands highlights the construction material of the different bridges. Data source: Yoram Reich and Steven J. Fenves, via the UCI Machine Learning Repository (Dua and Karra Taniskidou 2017).



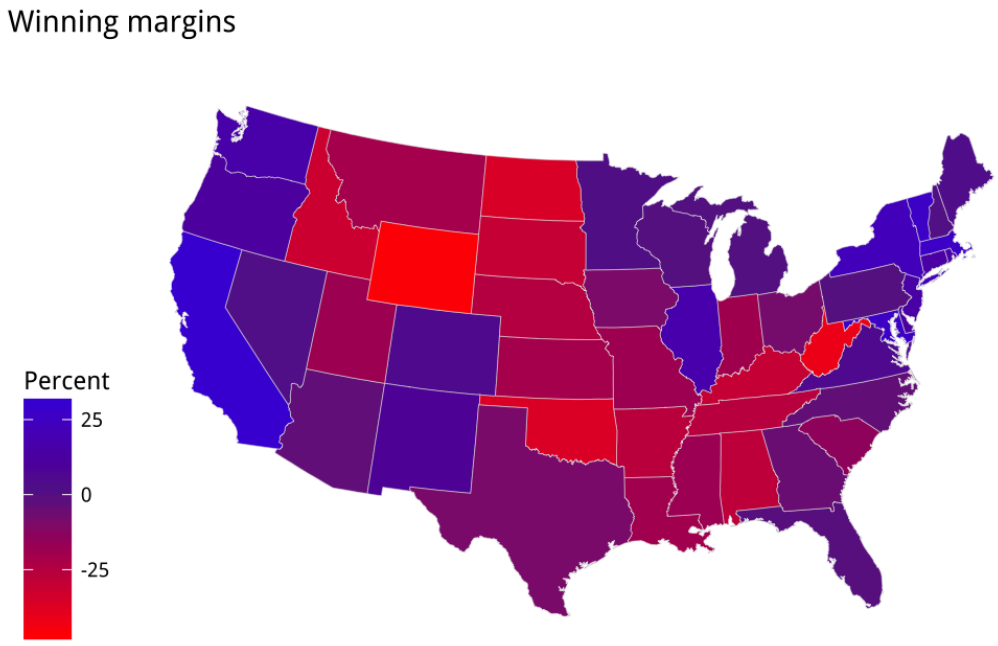
7. Give an example of a map-based visualization for continuous variable and discrete variable.

1) Continuous Variable

Example:

Percentage of wining margins in USA by state.

So, in this example percentage is shown in color.. We can represent blue color for +25 % and red for -25%.



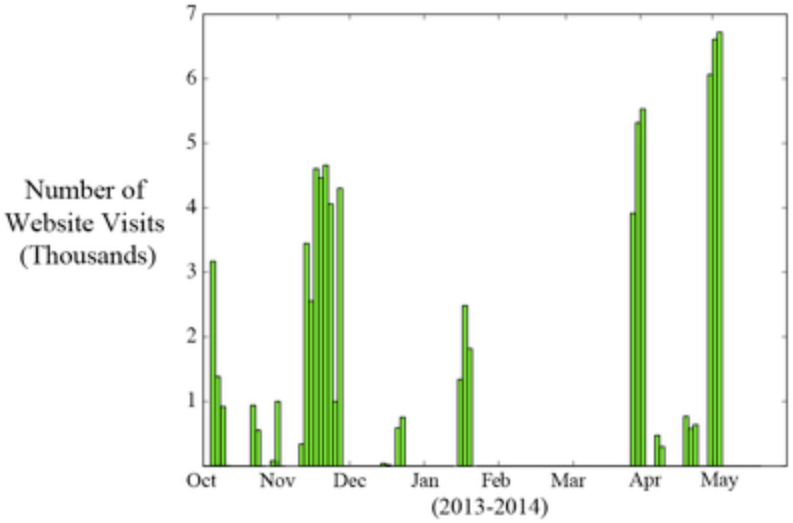
2) Discrete Variable

Example 1:

The rain percentage by states in USA for a particular year. The different percentages can be showed by different colors. So, Different states will be shown in different colors as per the rain percentage of the state.

Example 2:

Number of website visits in thousands for different months.


Number of
Website Visits
(Thousands)
7
6
5
4
3
2
1
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
(2013-2014)