

Rayanna Harduarsingh

September 30th, 2020

Module 6 Notes

IST 687

Chapter 12: Pictures Versus Numbers

Module 6: Introduction to Visualization

- “the communication of information using graphical representations.”
- Information visualization- the use of interactive visual representations of abstract data
 - It provides compact graphical presentations and user interfaces for interactively manipulating large numbers of items

Six key components to create a visualization:

- **Color:** The color of each symbol in the scatter plot. Color is the most common and was mentioned in our grid of numbers example. Note that some people are color-blind, so the use of color needs to take this into account.
- **Size:** The size of each symbol in the scatter plot.
- **Texture:** The shape of the symbol and whether the symbol is a solid color or a pattern.
- **Proximity:** The location of the symbol on the X-axis and the Y-axis.
- **Annotation:** Whether we label our scatter plots.
- **Interactivity:** Selecting one or more symbols, or perhaps zooming into a subset of the graph.

Ten Principles to create a visualization:

- Simplicity
- Encoding
 - Something that is bigger should be encoded as something longer/thicker/bigger as compared to something that is smaller.
- Patterns versus details
- Ranges
- Transformations
- Density
- Connections
- Aggregates

- Comparison
- Color

Sci Viz vs. Info Viz:

- **Visualization** is converting raw data to a form that is viewable and understandable to humans.
- **Scientific Visualization** is something with geometry associated with it and visualizing that data. Specifically concerned with data that has a well-defined representation in 2D or 3D space

Why Visualize?

- Vision is ‘highest bandwidth’ sense
 - Fast and parallel
 - Preattentive processing
- Eye trained for pattern recognition
 - Scanning
 - Recognizing
 - Remembering images

Using ggplot2

- Grammar of graphics
- A plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts.
- Provides a powerful model of graphics that makes it easy to produce complex multi-layered graphics.
- **Histogram being built in R:**

```
> g <- ggplot(dfStates, aes(x=july11pop))  
> g <- g + geom_histogram(binwidth=5000000,  
+   color="black", fill="white")  
> g <- g + ggtitle("states population histogram")  
> g
```

- **Seeing the distribution within a list of numbers to create a box plot in R:**

- `> ggplot(dfStates,aes(x=factor(0),july11pop)) +`
- `+ geom_boxplot()`

<code>hist()</code>	Generates a histogram using R's base graphics.
<code>barplot()</code>	Generates a bar chart using R's base graphics.
<code>ggplot()</code>	Starts the creation of a ggplot, defining the dataframe to be used.
Functions to Add Layers to a ggplot	
<code>ggtitle()</code>	Adds a title to the plot.
<code>geom_histogram()</code>	Creates a histogram.
<code>geom_boxplot()</code>	Creates a boxplot.
<code>geom_line()</code>	Creates a line chart.
<code>geom_col()</code>	Creates a bar chart.
<code>geom_point()</code>	Adds points to your plot.
<code>geom_text()</code>	Adds text to your plot.
<code>coord_flip()</code>	Rotates the chart by 90 degrees.
<code>theme()</code>	Refines the visual look of the chart (e.g., text).
<code>format()</code>	Enables a number to be printed in a pretty, easy-to-read format.
<code>scale_color_continuous()</code>	Defines a color range for mapping data to a color.