

## Lecture Notes- 6

### **Visualization:**

Information visualization is the use of interactive visual representations of abstract data which provides compact graphical presentations and user interfaces for interactively manipulating large number of items. It is concerned with data that does not have a well-defined representation in 2D or 3D space.

Information visualization is converting raw data to a form that is viewable and understandable to humans.

Scientific visualization is specifically concerned with a well-defined representation in 2D or 3D space.

Visualization makes it easier to understand the data and recognizes patterns. Components of visualization include color, size, texture, proximity, annotation and interactivity.

### **What is misleading or confusing about these two visualizations?**

The bar chart the y chart axis is too zoomed so misleads the audience as ratio at the axis should give a good perspective to the audience. The other chart was stacked on top so not a clear comparison so should be stacked next to each other and no information on the y axis so no specific analysis. It should be easily understandable, and the message should be clear to the audience.

### **Ten principles in Visualization:**

Simplicity- create a simplest graph to cover the information.

Encoding- consider things like encoding, width, color

Patterns vs Details- Focus on patterns or on visualizing

Ranges- select meaningful axis ranges

Transformations- Data transformations and aspect ratios should be used to emphasize the content

Show Density- Density differences apart help to understand the data more

Connections- use lines when connecting sequential data in time-series plots

Aggregates- aggregate larger data sets in meaningful ways

Comparison- keep axis ranges as similar as possible

Colors- select appropriate color scheme based on the type of data

### **What is good or bad about these charts (stacked human and spiral piechart)?**

The stacked person chart was interesting with no proper comparison and the chart percentages do not add up to 100. The Ted talk spiral chart had no proper ordering and confusing and added not much value instead a simple pie chart could be used. It is important to know the audience, be clear and easy to read. The map chart provided clear information and the heatmap is easily understood in terms of being easier to interpret by people.

### **GGPLOT:**

Ggplot2 is a package used for visualizations which is comprised of three plots:  
Plot= Data + Aesthetics + Geometry

Data is data frame, aesthetics is used to indicate x and y variables and used to control color, size, shape, etc. Geometry defines the type of graphics. It is known as ggplot as it is the “grammar of graphics”.

### **Example:**

```
ggplot(dfStates, aes(x=july11pop)) + geom_histogram(bins=30)
#dfStates is the vector, x axis is column and geom is histogram
ggplot(dfStates, aes(x=july11pop)) + geom_histogram(binwidth=1000000, color="black",
fill="white")
#configure color and fill
ggplot(dfStates, aes(x=factor(0), july10pop))+ geom_boxplot()
#create a box plot
```

### **Code 1:**

```
g<- ggplot(travel.df, aes(x=datOfWeek, group=week, color=week))+geom_line(aes(y=time))
#aes is x axis and group by week and color code by week and store in variable
g<- g+ geom_point(y=time, color="black", size=4, shape=21, fill="white")
```

```
#adds data points and colors by size and shape
g <- g+ ylab("time to NYC (in hours)") + ggtitle("compare weekly times")
#add table and ylabel
```

GGplot gives more control than base R package. Bin width is logical operation with bucket size and bins is number of bins. It allows to add more layers to the visualization.

### Code 2:

```
install.packages("ggplot2")
library(ggplot2)
#install ggplot2 package
df <- data.frame(dayOfWeek, time, week)
df
#create dataframe
g <- ggplot(df, aes(x=dayOfWeek, group=week, color=week)) +geom_line(aes(y=time))
g <-g +ylab("time to NYC (in hours)") + ggtitle("comparing weekly times")
#create a line chart, group by week and color by week and add yaxis label and title.
```

### Code 3:

```
mtc<- mtcars
ggplot(mtcars, aes(x=factor(0), mpg)) + geom_boxplot()
#create a box plot and factor 0 so that is is lumped together
ggplot(mtcars, aes=(group=cyl, x=cyl)) + geom_boxplot() + coord_flip()
#coord_flip() is rotate
ggplot (mtcars, aes=(x=cyl)) + geom_bar()
#create a barchar
hist(mtc$cyl)
#create a histogram
cars.name <- rownames(mtc)
#retrieves row names
car.names
```

```
g <- ggplot(mtcars, aes(x=car.names, y=wt)) + geom_bar(stat="identity")  
g  
#create a barchart  
g <- g + theme(axis.text, x=element_text(angle=90, hjust=1))  
#for making reading axis text easier  
g+ ggtitle("my car weight car")  
# add a title
```