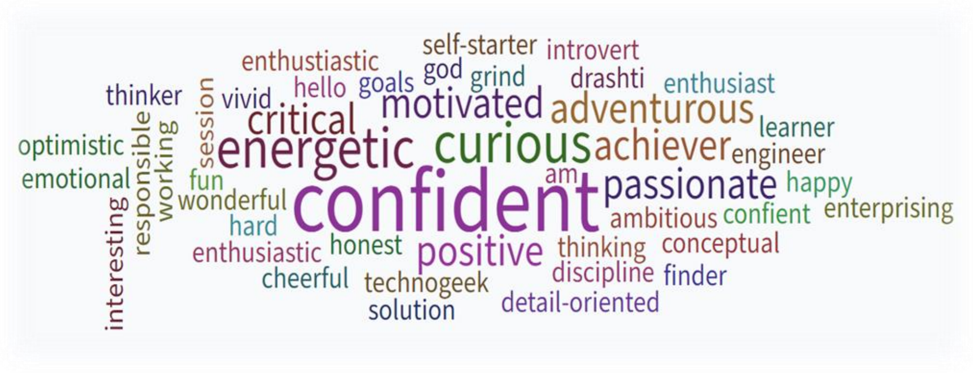
Northeastern University
Toronto

ALY 6000
Introduction to Data Analytics

Mohammad (**Shafiqul**) Islam, PhD., P.Eng.
Email:m.islam@northeastern.edu



Northeastern
University

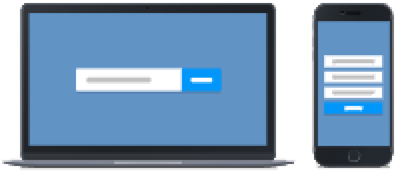
Agenda

- Module 1 Review
- Module 2
- Module Project
- Summary

- I will use this deck as a reference for this class

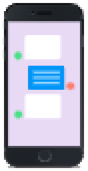
Pulse Check

Join by Web



1. Go to **PollEv.com**
2. Enter **MISLAM933**
3. Respond to activity

Join by Text



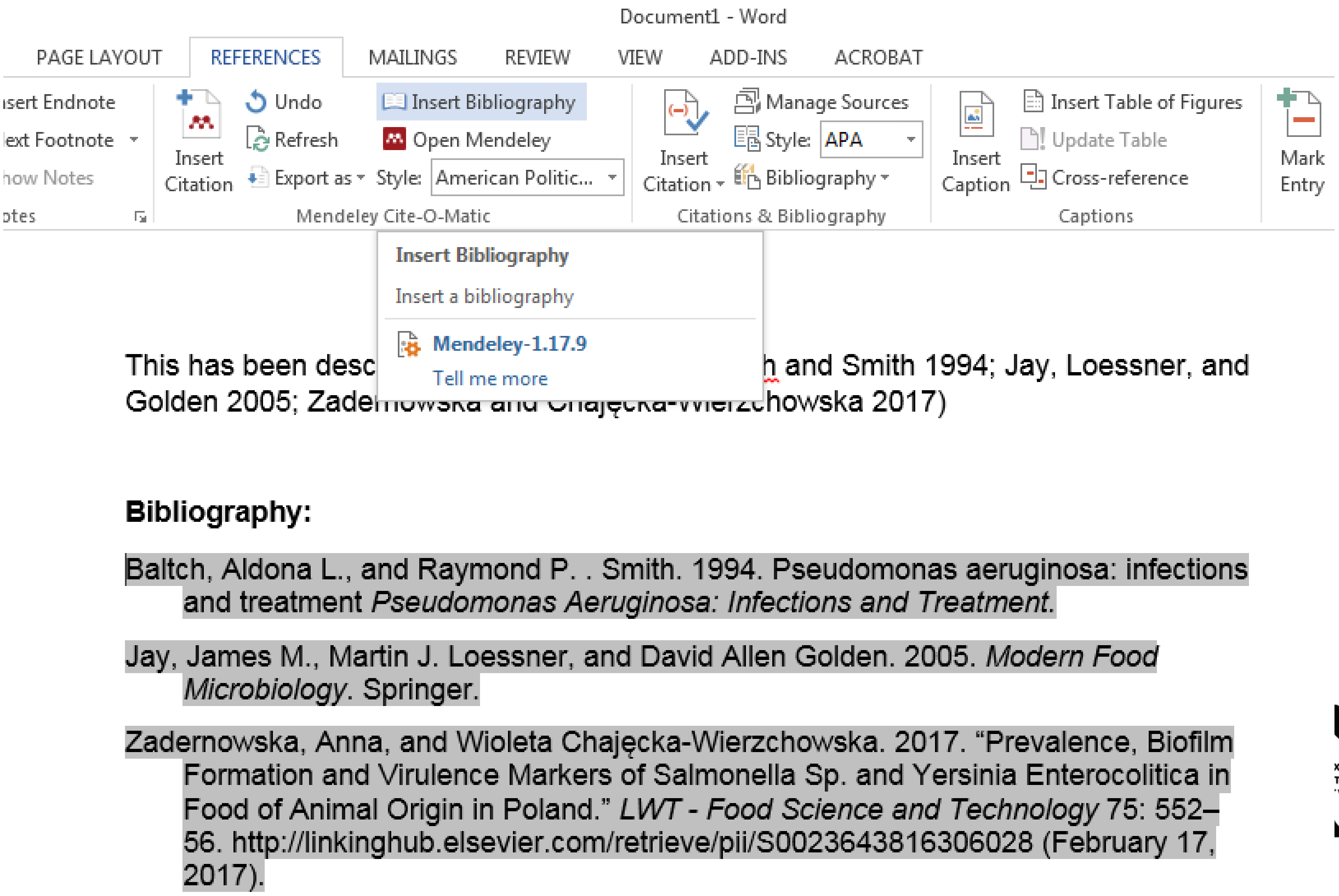
1. Text **MISLAM933** to **37607**
2. Text in your message

Module 1 Review

Key Issues:

- Git/Github: How to commit your code?
- R Code:
  - Data Reading Warning!
  - Where to put my code in the report?
- Reporting:
  - What to write?
  - How to write APA references?
- Submission Process
- Submission Status
- Grading
- Profile Pictures
- Joining to Zoom

Inserting Bibliography Using Mendeley



This has been desc h and Smith 1994; Jay, Loessner, and Golden 2005; Zadernowska and Chajęcka-Wierzchowska 2017)

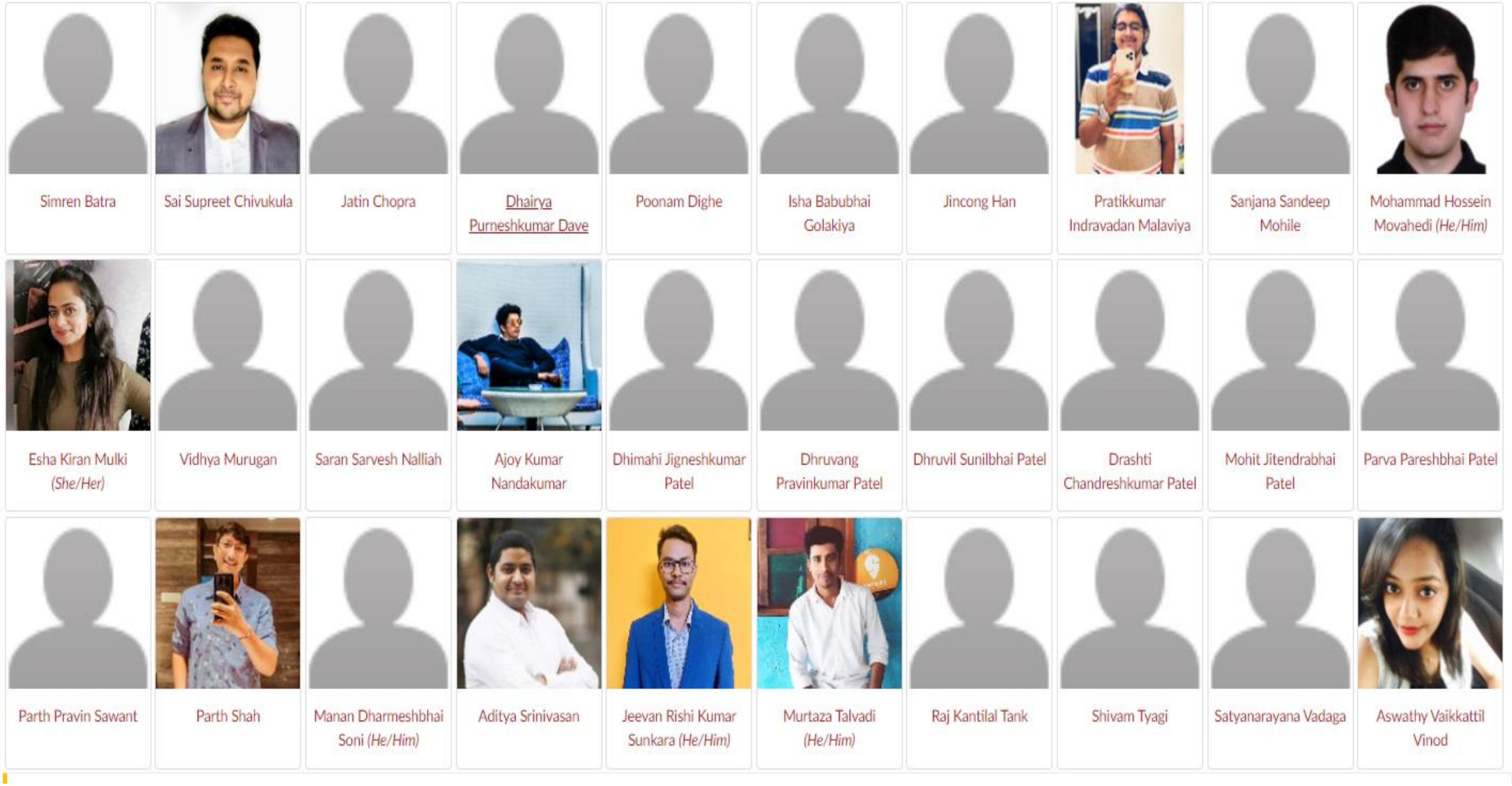
Bibliography:

Baltch, Aldona L., and Raymond P. . Smith. 1994. Pseudomonas aeruginosa: infections and treatment *Pseudomonas Aeruginosa: Infections and Treatment.*

Jay, James M., Martin J. Loessner, and David Allen Golden. 2005. *Modern Food Microbiology*. Springer.

Zadernowska, Anna, and Wioleta Chajęcka-Wierzchowska. 2017. “Prevalence, Biofilm Formation and Virulence Markers of Salmonella Sp. and Yersinia Enterocolitica in Food of Animal Origin in Poland.” *LWT - Food Science and Technology* 75: 552–56. http://linkinghub.elsevier.com/retrieve/pii/S0023643816306028 (February 17, 2017).

Profile Pictures



Simren Batra

Sai Supreet Chivukula

Jatin Chopra

Dhairya Purneshkumar Dave

Poonam Dighe

Isha Babubhai Golakiya

Jincong Han

Pratikkumar Indravadan Malaviya

Sanjana Sandeep Mohile

Mohammad Hossein Movahedi *(He/Him)*

Esha Kiran Mulki *(She/Her)*

Vidhya Murugan

Saran Sarvesh Nalliah

Ajoy Kumar Nandakumar

Dhimahi Jigneshkumar Patel

Dhruvang Pravinkumar Patel

Dhruvil Sunilbhai Patel

Drashti Chandreshkumar Patel

Mohit Jitendrabhai Patel

Parva Pareshbhai Patel

Parth Pravin Sawant

Parth Shah

Manan Dharmeshbhai Soni *(He/Him)*

Aditya Srinivasan

Jeevan Rishi Kumar Sunkara *(He/Him)*

Murtaza Talvadi *(He/Him)*

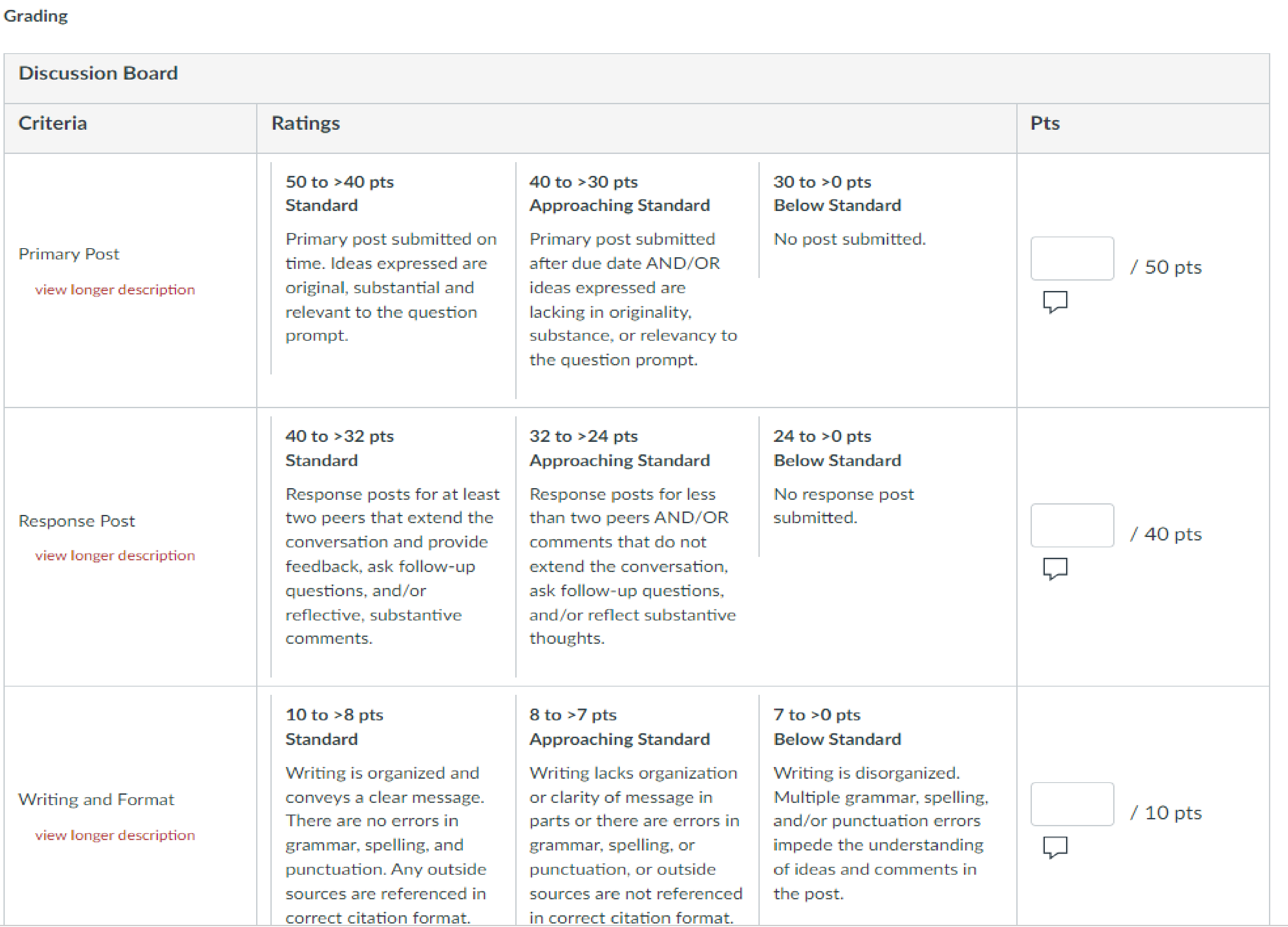
Raj Kantilal Tank

Shivam Tyagi

Satyanarayana Vadaga

Aswathy Vaikkattil Vinod

Discussion Grading

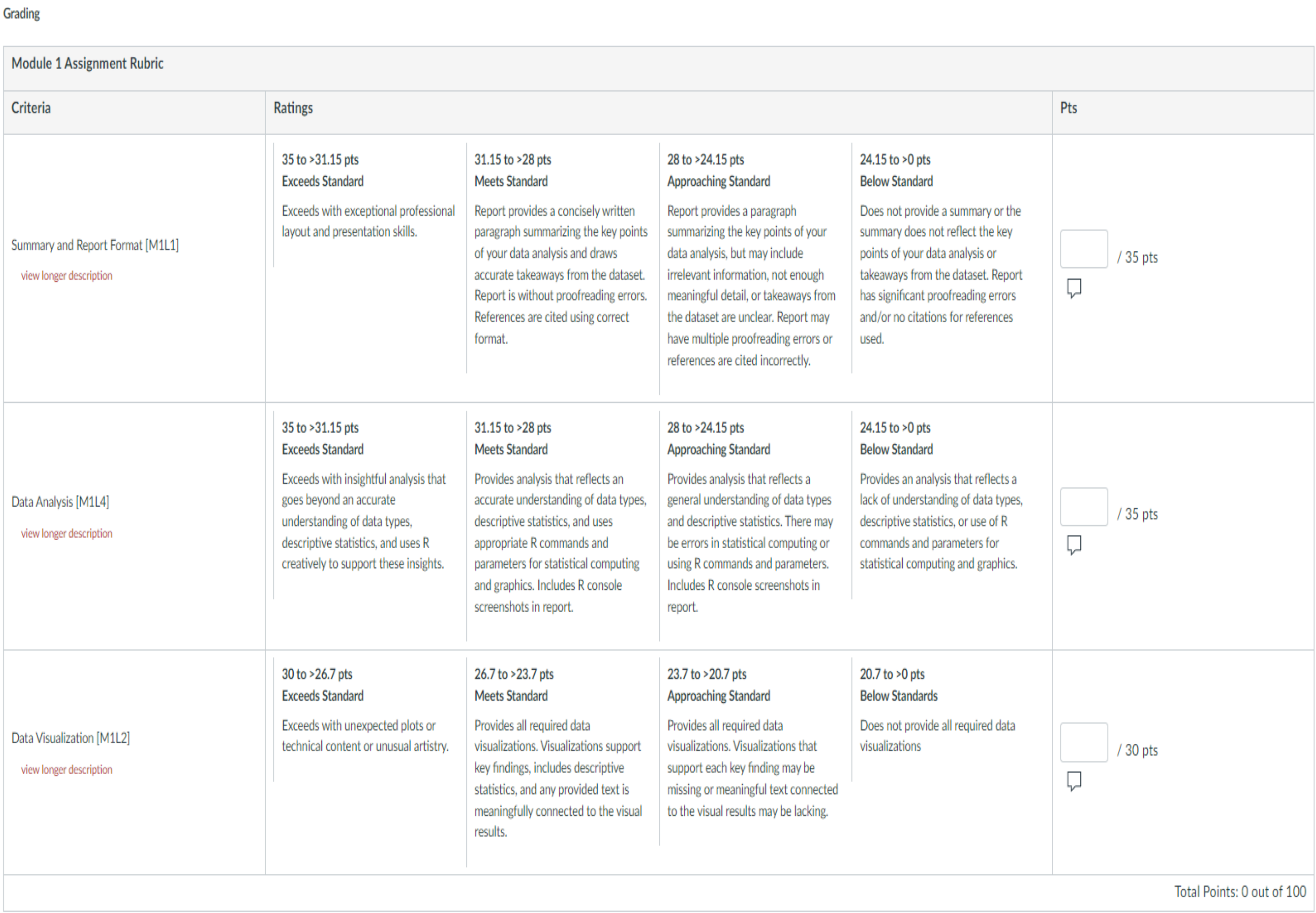


Grading

Discussion Board

Criteria	Ratings			Pts
Primary Post view longer description	**50 to >40 pts** **Standard** Primary post submitted on time. Ideas expressed are original, substantial and relevant to the question prompt.	**40 to >30 pts** **Approaching Standard** Primary post submitted after due date AND/OR ideas expressed are lacking in originality, substance, or relevancy to the question prompt.	**30 to >0 pts** **Below Standard** No post submitted.	□ / 50 pts
Response Post view longer description	**40 to >32 pts** **Standard** Response posts for at least two peers that extend the conversation and provide feedback, ask follow-up questions, and/or reflective, substantive comments.	**32 to >24 pts** **Approaching Standard** Response posts for less than two peers AND/OR comments that do not extend the conversation, ask follow-up questions, and/or reflect substantive thoughts.	**24 to >0 pts** **Below Standard** No response post submitted.	□ / 40 pts
Writing and Format view longer description	**10 to >8 pts** **Standard** Writing is organized and conveys a clear message. There are no errors in grammar, spelling, and punctuation. Any outside sources are referenced in correct citation format.	**8 to >7 pts** **Approaching Standard** Writing lacks organization or clarity of message in parts or there are errors in grammar, spelling, or punctuation, or outside sources are not referenced in correct citation format.	**7 to >0 pts** **Below Standard** Writing is disorganized. Multiple grammar, spelling, and/or punctuation errors impede the understanding of ideas and comments in the post.	□ / 10 pts

Report Grading



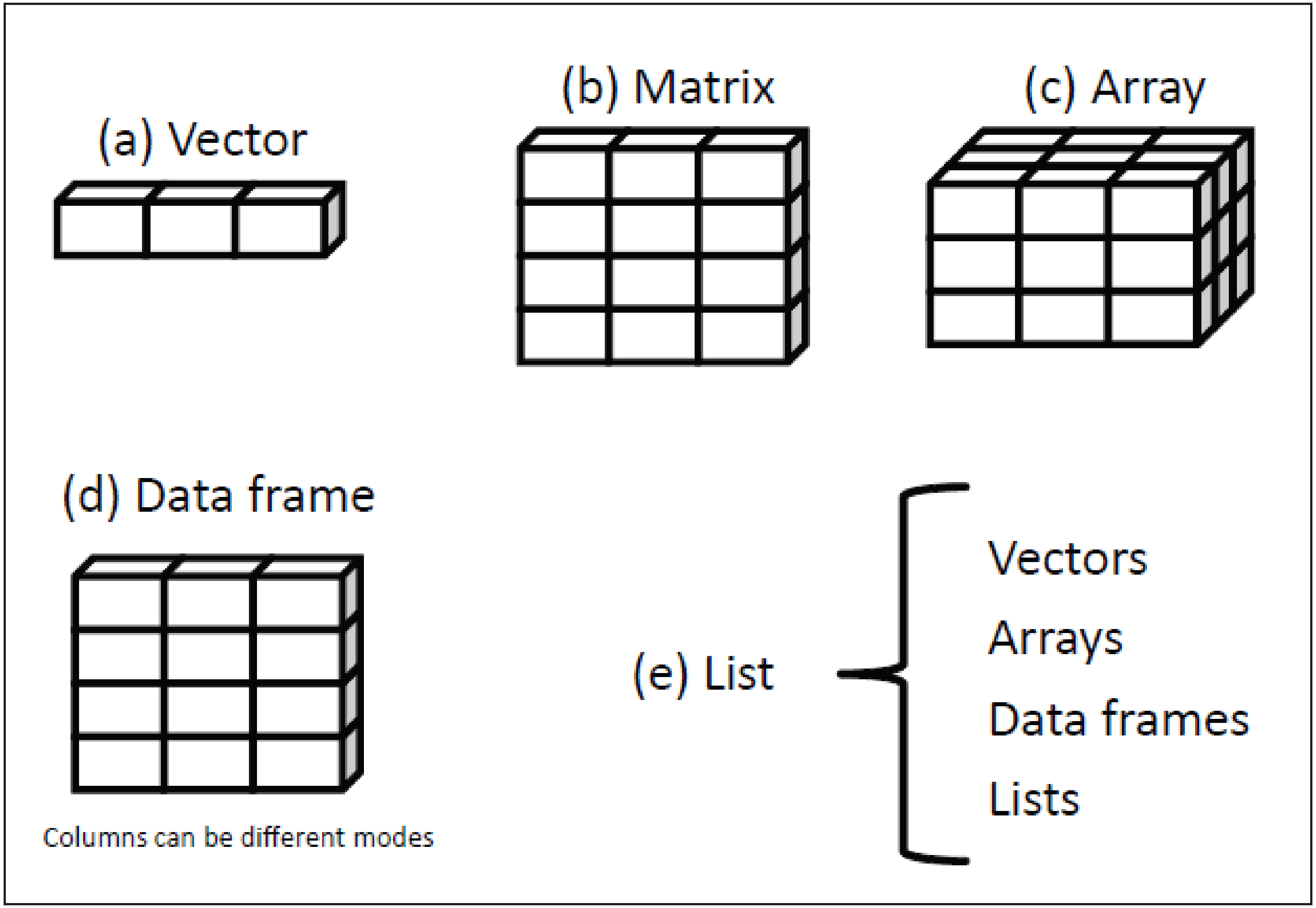
Grading

Module 1 Assignment Rubric

Criteria	Ratings				Pts
Summary and Report Format [M1L1] view longer description	35 to >31.15 pts Exceeds Standard Exceeds with exceptional professional layout and presentation skills.	31.15 to >28 pts Meets Standard Report provides a concisely written paragraph summarizing the key points of your data analysis and draws accurate takeaways from the dataset. Report is without proofreading errors. References are cited using correct format.	28 to >24.15 pts Approaching Standard Report provides a paragraph summarizing the key points of your data analysis, but may include irrelevant information, not enough meaningful detail, or takeaways from the dataset are unclear. Report may have multiple proofreading errors or references are cited incorrectly.	24.15 to >0 pts Below Standard Does not provide a summary or the summary does not reflect the key points of your data analysis or takeaways from the dataset. Report has significant proofreading errors and/or no citations for references used.	/ 35 pts
Data Analysis [M1L4] view longer description	35 to >31.15 pts Exceeds Standard Exceeds with insightful analysis that goes beyond an accurate understanding of data types, descriptive statistics, and uses R creatively to support these insights.	31.15 to >28 pts Meets Standard Provides analysis that reflects an accurate understanding of data types, descriptive statistics, and uses appropriate R commands and parameters for statistical computing and graphics. Includes R console screenshots in report.	28 to >24.15 pts Approaching Standard Provides analysis that reflects a general understanding of data types and descriptive statistics. There may be errors in statistical computing or using R commands and parameters. Includes R console screenshots in report.	24.15 to >0 pts Below Standard Provides an analysis that reflects a lack of understanding of data types, descriptive statistics, or use of R commands and parameters for statistical computing and graphics.	/ 35 pts
Data Visualization [M1L2] view longer description	30 to >26.7 pts Exceeds Standard Exceeds with unexpected plots or technical content or unusual artistry.	26.7 to >23.7 pts Meets Standard Provides all required data visualizations. Visualizations support key findings, includes descriptive statistics, and any provided text is meaningfully connected to the visual results.	23.7 to >20.7 pts Approaching Standard Provides all required data visualizations. Visualizations that support each key finding may be missing or meaningful text connected to the visual results may be lacking.	20.7 to >0 pts Below Standards Does not provide all required data visualizations	/ 30 pts

Total Points: 0 out of 100

Data Structures



Columns can be different modes

Vectors and Matrices

Vector:

Vectors are one-dimensional arrays that can hold numeric data, character data, or logical data.

```
a <- c(1, 2, 5, 3, 6, -2, 4)
```

Matrix:

A matrix is a two-dimensional array where each element has the same mode (numeric, character, or logical).

```
myymatrix <- matrix(vector, nrow=number_of_rows,
ncol=number_of_columns,byrow=logical_value,
dimnames=list(char_vector_rownames,
char_vector_colnames))
```

```
y <- matrix(1:20, nrow=5, ncol=4)
```

Arrays

Arrays are similar to matrices but can have more than two dimensions. They’re created with an array function of the following form:

```
›myarray <- array(vector, dimensions, dimnames)


›dim1 <- c("A1", "A2")
›dim2 <- c("B1", "B2", "B3")
›dim3 <- c("C1", "C2", "C3", "C4")
› z <- array(1:24, c(2, 3, 4),
 dimnames=list(dim1, dim2, dim3))
```

Dataframes

A data frame is more general than a matrix in that different columns can contain different modes of data (numeric, character, etc.). It's similar to the datasets you'd typically see in SAS, SPSS, and Stata. Data frames are the most common data structure you'll deal with in R.

Arrays are similar to matrices but can have more than two dimensions. They're created with an array function of the following form:

```
>mydata <- data.frame(col1, col2, col3,…)

>patientID <- c(1, 2, 3, 4)
>age <- c(25, 34, 28, 52)
>diabetes <- c("Type1", "Type2", "Type1", "Type1")
>status <- c("Poor", "Improved", "Excellent", "Poor")
>patientdata <- data.frame(patientID, age, diabetes,
 status)
```

Factors

- Factors represent categorical data.
- Can be ordered or unordered.
- Factors are stored as integers, and have characters/ labels associated with these unique integers

Example:

- `status <- c("Poor", "Improved", "Excellent", "Poor")`

List

- Lists are the R objects which contain elements of different types like − numbers, strings, vectors and another list inside it. A list can also contain a matrix or a function as its elements. List is created using list() function.

Example:

-
  ```
  list_data <- list("Red", "Green", c(21,32,11), TRUE, 51.23, 119.1)
  ```

-
  ```
  print(list_data)
  ```

Module 2:
Frequency Distribution, Data Description and Graphing

Module Overview

- As data analysts, it’s important to be able to describe the data that you are working with, both numerically and graphically. Numerically describing your data includes determining the center, variability and shape of all the data points together. Graphically describing data means being able to choose an appropriate type of graph and then use that graph to display the numerical data clearly.

- In this module we will talk about various measures that help describe your data numerically and ways to display that information.

Learning Objectives

By the end of this module, you should be able to:

- Calculate basic descriptive statistics to describe a set of data
- Create various types of graph based on data provided
- Use descriptive statistics and graphs to describe and explain data
- Use R to visualize data

Task List

- Statistics (Bluman textbook - Chapters 2 and 3)
- Complete primary Discussion post by Friday
- Learning to use R Practice (do assigned reading, watch instructor videos, complete R practice tasks)
  - Kabacoff textbook - Chapter 3
- Complete Module 2 Project

Common Notations

N: number of elements in a population

μ: population mean

σ: population standard deviation

σ^2: population variance

n: number of elements in a sample

$\overline{X}$: sample mean

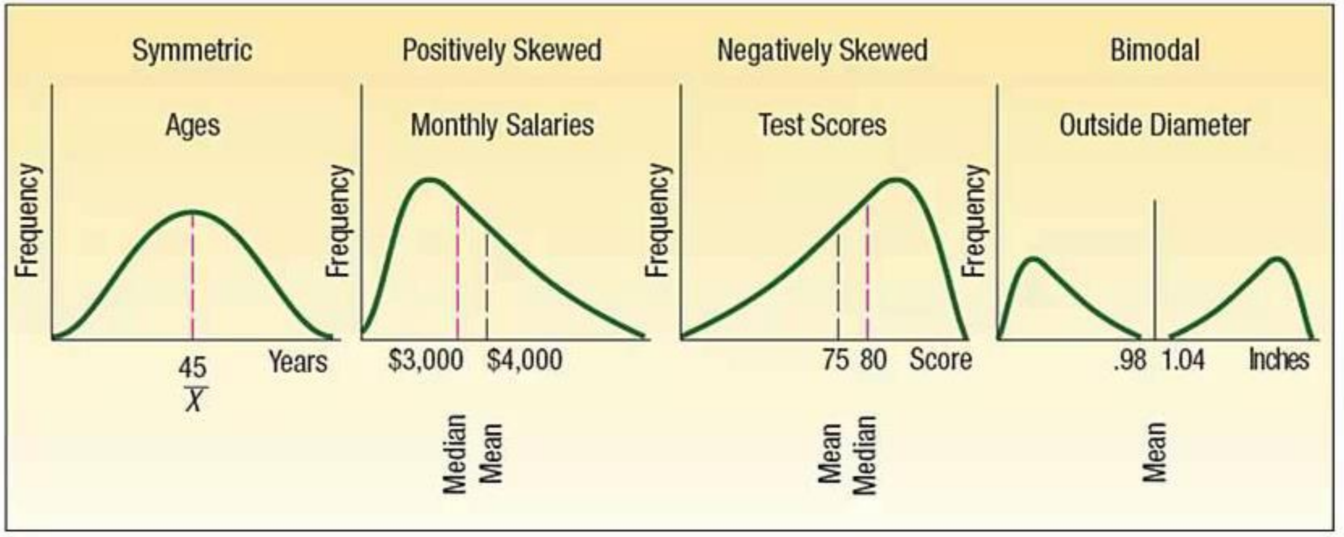
s: sample standard deviation

s^2 : sample variance

Characteristics of Numerical Data

When describing numerical data, there are three characteristics we want to capture:

- the center,
- the variability or dispersion, and
- the shape

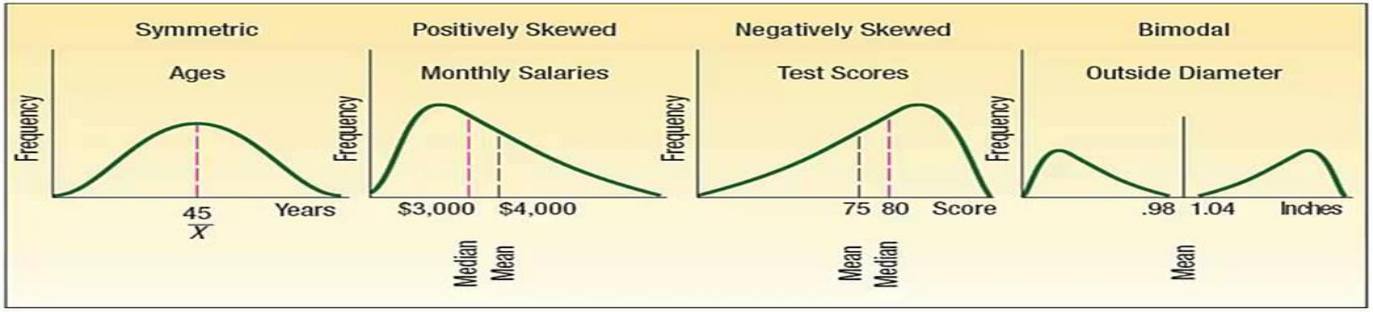


Characteristics of Numerical Data

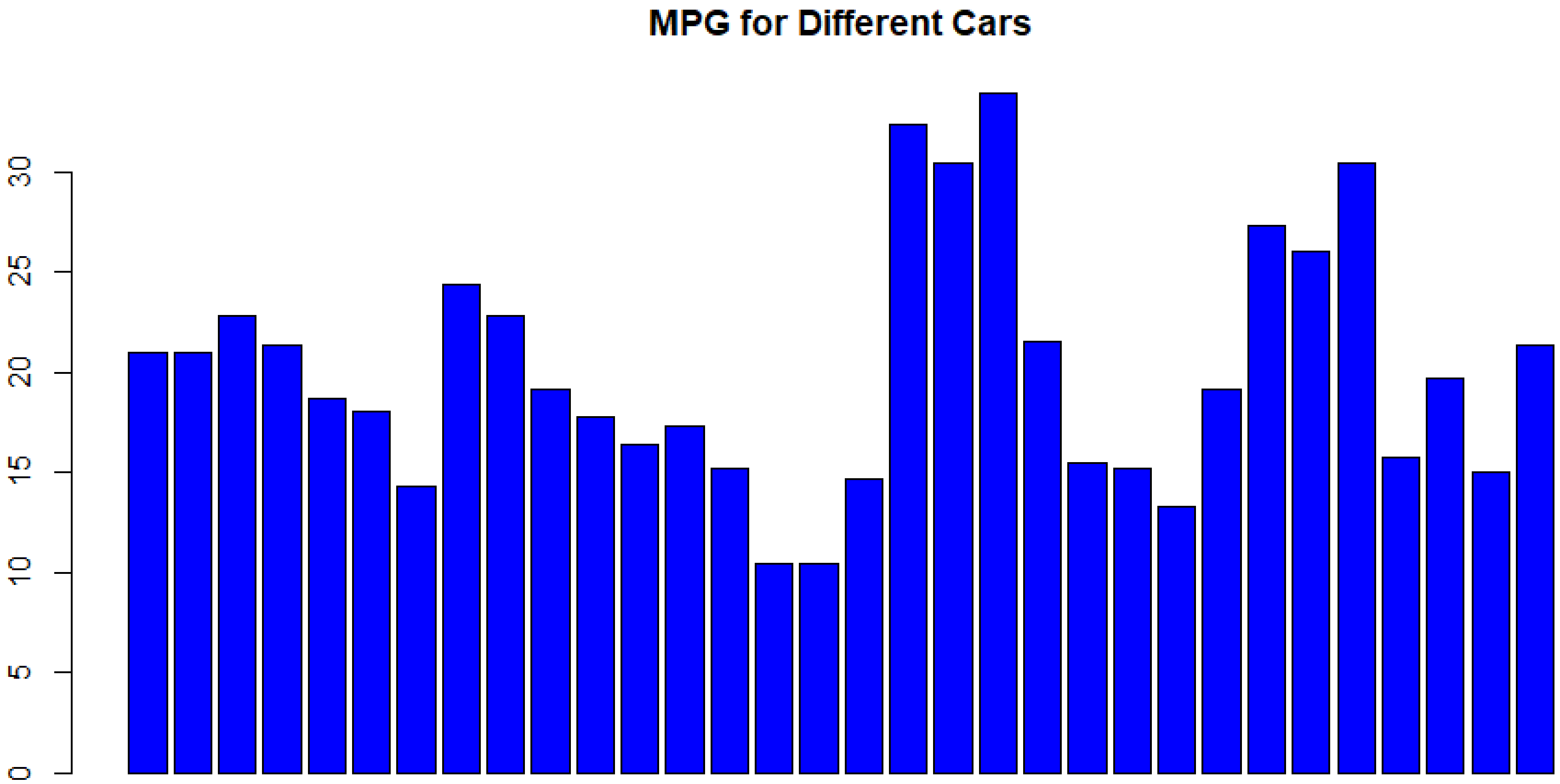
- The measure of central tendency of the distribution is a measure of where "the middle" is.

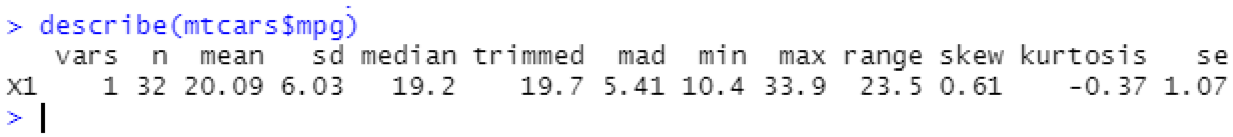
- Another important metric is a measure of the dispersion, or of how spread out the data is. This is a measure of how far away the data is from the center.

- The shape of the data is visually shown in the distribution curve. We consider how pointy the curve is near its peak, and the skewness of the curve or how asymmetric or lopsided the curve is.
- The kurtosis and skewness indicate the pointy-ness and which way the data is leaning.



Characteristics of Numerical Data

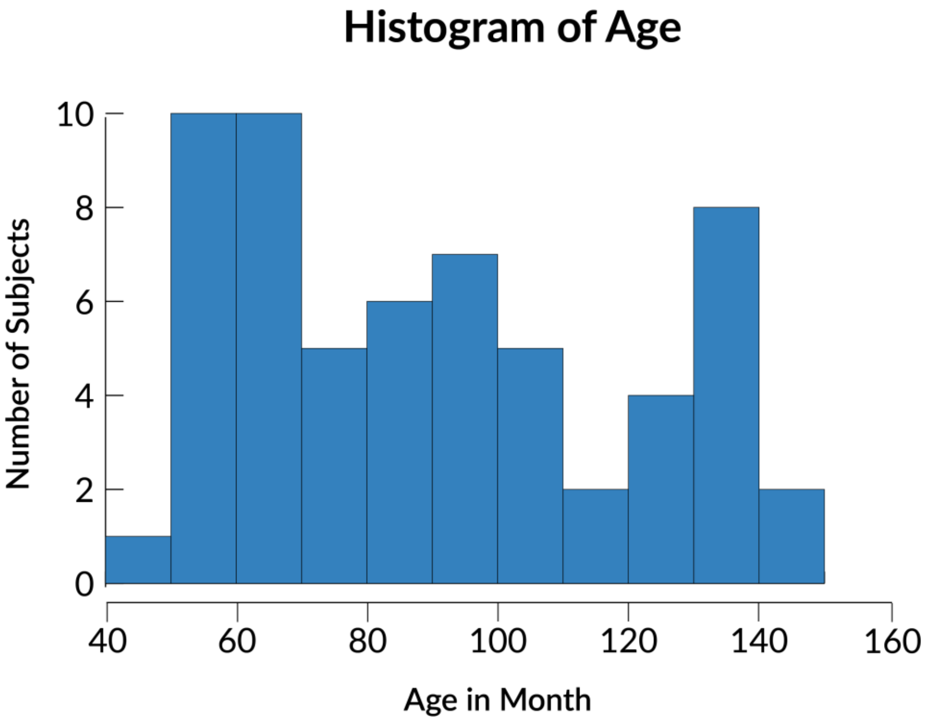




```
> describe(mtcars$mpg)
   vars  n  mean   sd median trimmed  mad  min  max range skew kurtosis   se
X1    1 32 20.09 6.03   19.2    19.7 5.41 10.4 33.9  23.5 0.61    -0.37 1.07
> |
```

Characteristics of Numerical Data

Mean	90.42
Standard Error	3.90
Median	84
Mode	60
Standard Deviation	30.23
Sample Variance	913.84
Kurtosis	-1.18
Skewness	0.39
Range	95
Minimum	48
Maximum	143
Sum	5425
Count	60



Mean: Measures of Central Tendency

The average, or the mean, is one of the most common indicators of central tendency, or of the central location of the data. It’s the sum of all the values divided by the number of values in the data set.

- It is the value to expect, on average and in the long run

- Other types of means are geometric and harmonic, these are outside the scope of this class.

The Mean

- For a population of size *N*, the **population mean** (μ) is defined as $\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1+X_2+\cdots+X_N}{N}$
  - It is the value to expect, on average and in the long run

- For a sample of size *n*, the **sample mean** ($\bar{x}$) is defined as $\mu = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1+X_2+\cdots+X_n}{n}$

Weighted Mean

- In some situations, the measurements in a set of data are different weights or are of different duplicities. In these types of situations, the mean is said to be a weighted mean.

- Calculate weighted mean as $\frac{\sum w_i x_i}{\sum x_i}$ where w_i is the weight assigned to the i^{th} measurement x_i

Median

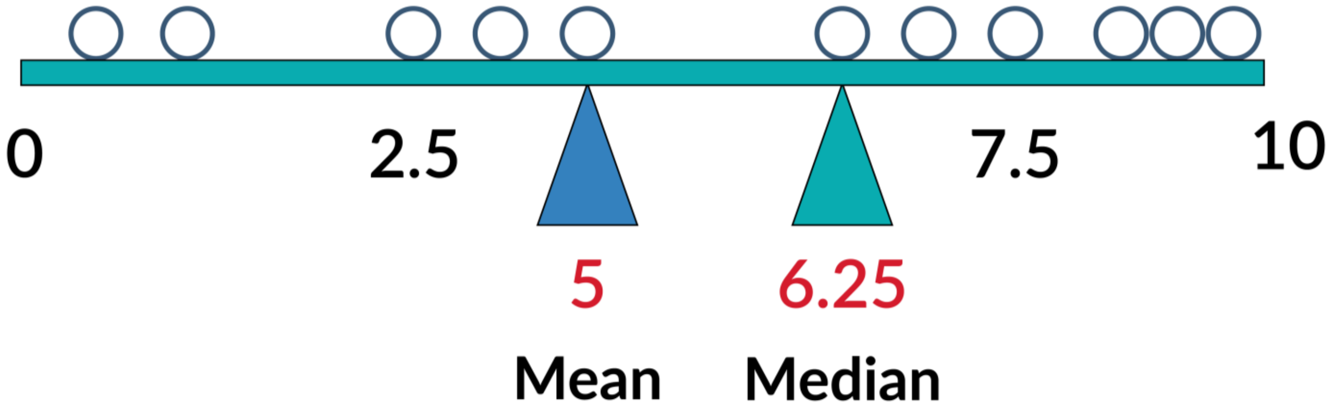
- We have talked about the mean or average being one measure of the central tendency of a data set. The median is another common measure of the central tendency of a data set. It is simply the middle value of the sorted data set.
- If there are an odd number of values, the median is just the middle value of the sorted data set. If there are an even number of values, the median is the mean of the two middle values of the sorted data set. This means you add the two middle values together and divide by two.

- For example: (45, 49, 50, 53, 60, 62, 63, 65, 66, 67, 69, 71, 73, 74, 74, 78, 81, 85, 87, 100)

- Median: 68

Mean and Median

- Both the mean and the median are useful in describing the central measure of a data set, but they are not always the same.
- The two values are usually different, and it's up to us to determine the better measure to use.



Mode

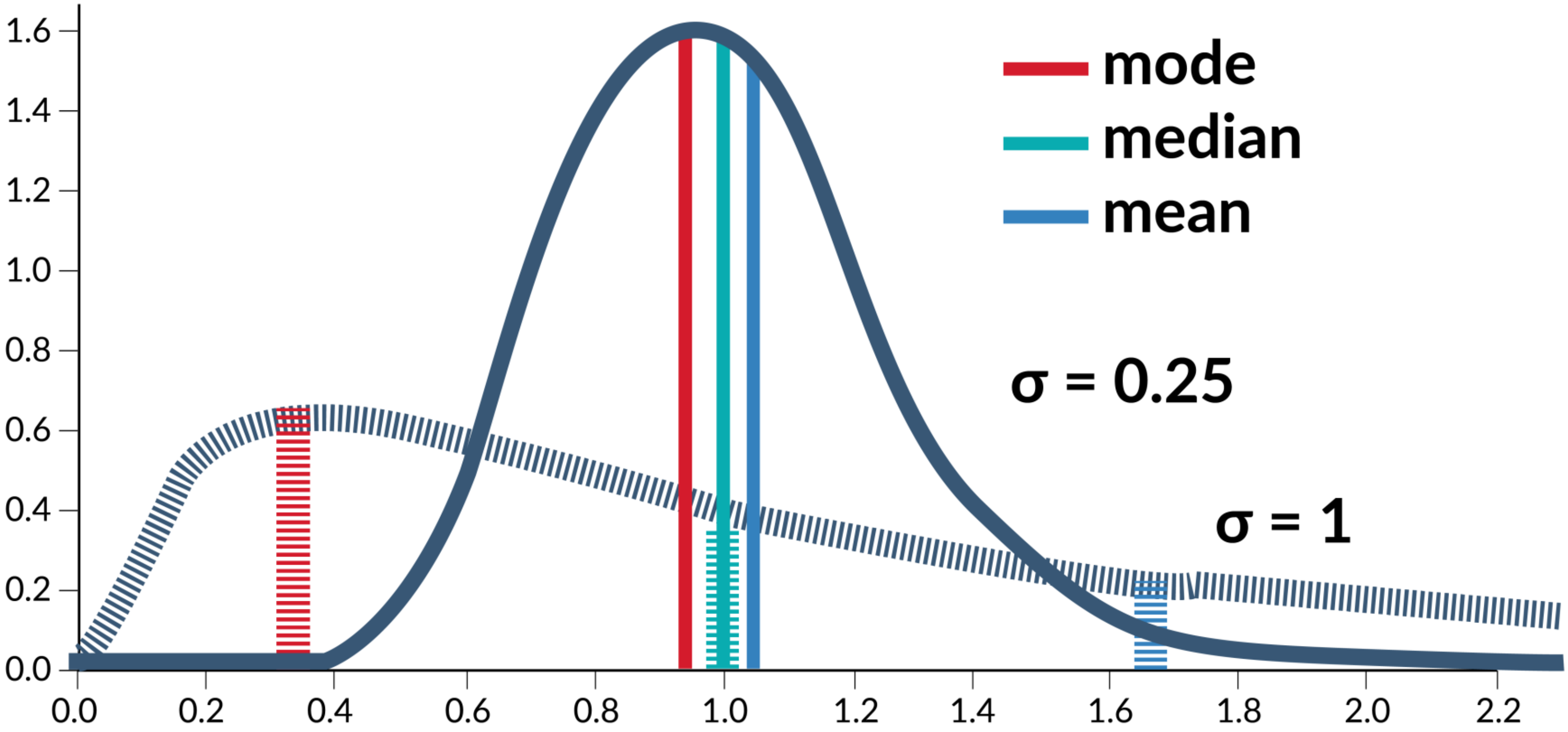
- Mode is the most frequently occurring value.

- The mode is most useful when we have a discrete variable, and it may also be applied to nominal, categorical data. For example, if we're looking at hair color: It is reasonable to have a mode of brown hair.
- Example:

(45, 49, 50, 53, 60, 62, 63, 65, 66, 67, 69, 71, 73, 74, 74, 78, 81, 85, 87, 100)

- In this data set, the mode is 74.

Mean, Median, and Mode



Range

- It is the difference between the maximum and minimum values in the data set.
- Largest minus smallest
- Measures the interval spanned by all the data
- A larger range usually (but not always) indicates a large spread or deviation in the values of the data set.
- Example:

{73, 66, 69, 67, 49, 60, 81, 71, 78, 62, 53, 87, 74, 65, 74, 50, 85, 45, 63, 100}

- Range: (45,100)

Variance

- The *variance* is the average of the squared deviations of the individual measurements from the mean
- For a population of size *N*, the population variance is
- $\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}$
- For a sample of size *n*, the sample variance is $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_{n-1} - \bar{x})^2}{n-1}$

Standard Deviation

- This quantity is the square root of the variance, and is denoted by σ for a population, and by s for a sample.
- The standard deviation is the most widely used measure of dispersion. An intuitive way to think of these values is that they measure the deviation from the mean of the data set.

- Population standard deviation is $\sigma = \sqrt{\sigma^2}$

- Sample standard deviation is s$= \sqrt{s^2}$

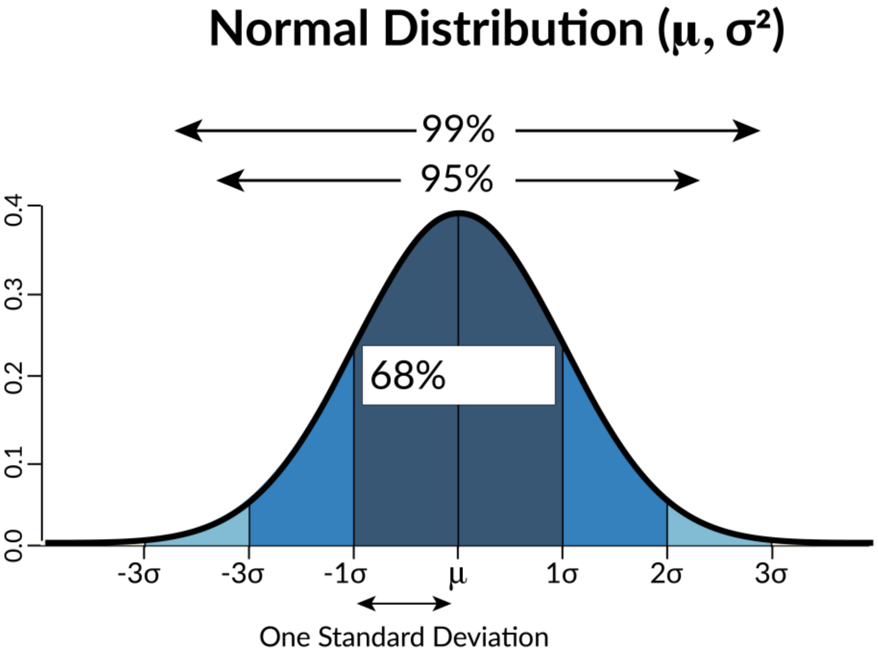
Class Exercise:

- Calculate followings for the given sample data:
  - Mean
  - Median
  - Mode
  - Variance and
  - Standard deviation

- Data: -2, 0, 1, 2, 5, 5, 6,10.

Normal Distribution

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.



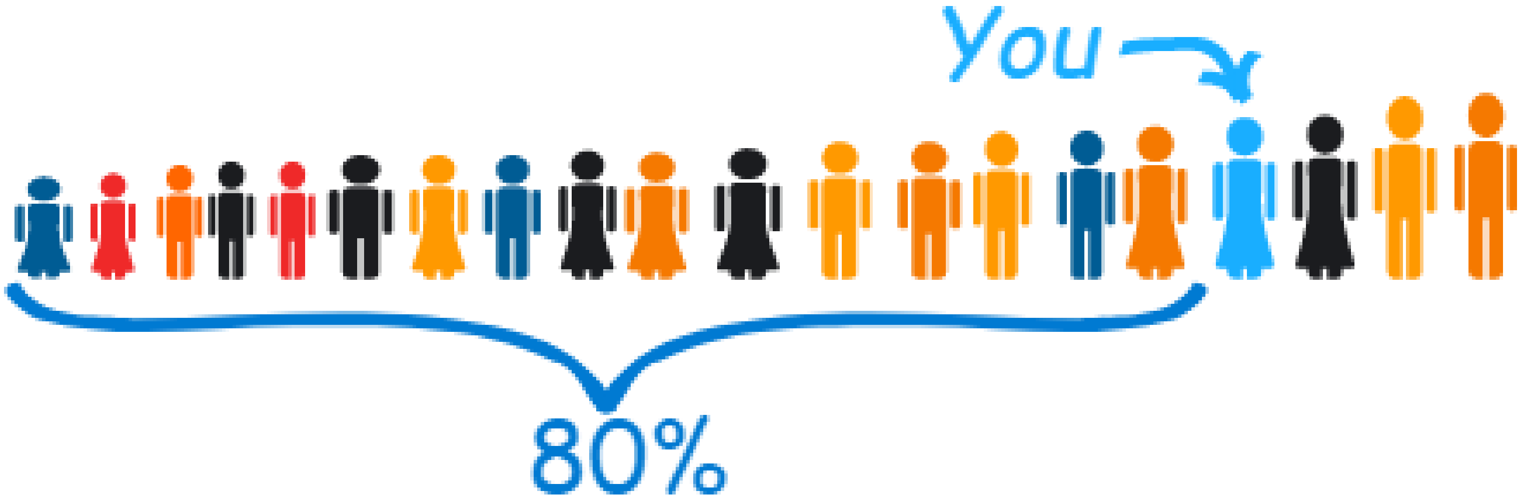
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Coefficient of Variation

- Coefficient of variation measures the size of the standard deviation relative to the size of the mean
- The coefficient of variation is simply the standard deviation divided by the mean.
- $\frac{Standard\ deviation}{Mean} \times 100\% = \frac{\sigma}{\bar{x}} \times 100\%$
- With this normalization of the standard deviation using the mean of the data, the coefficient of variation statistic can be used directly as a measure of dispersion across data sets.

Percentiles

The value below which a percentage of data falls.



That means you are at the 80th percentile.

If your height is 1.85m then "1.85m" is the 80th percentile height in that group.

Source: www.mathsisfun.com

Calculating Percentiles

- Sort the data.

- Using PC, the desired percentile. calculate the location of the value of interest, n_{PC} $$n_{PC} = \left(\frac{n+1}{100}\right) PC$$

- If n_{PC}, is an integer, the value at that location in the sorted data set is the value of the desired percentile.

- If n_{PC} is not an integer, use the two nearest values to n_{PC} in the sorted dataset to calculate n_{PC}.

Quartiles and Decile

- The **first quartile Q_1** is the 25th percentile
- The **second quartile** (median) is the 50th percentile
- The **third quartile Q_3** is the 75th percentile
- The **interquartile range IQR** is Q_3 - Q_1

- Decile analysis helps us to understand the relative importance of sections of the variable distribution. We order the data then divide it into 10 slots. These 10 groups of data are known as deciles.

Source: www.mathsisfun.com

Deciles

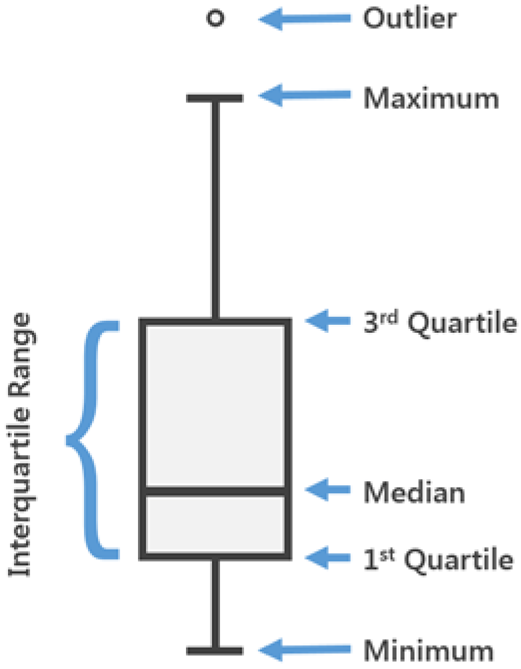
- Decile analysis helps us to understand the relative importance of sections of the variable distribution. We order the data then divide it into 10 slots. These 10 groups of data are known as deciles.

- Since the data is broken into equal cut points, we can summarize how much contribution each decile makes to the overall values.

- In the case of revenue, if each data point how much a particular customer spends, we can calculate how much the customers who are in the top decile spend and compare that to how much customers in the bottom decile spend. That gives us the relative importance of the customers to the total revenue.

Outliers

- Outliers are measurements that are very different from other measurements
- They are either much larger or much smaller than most of the other measurements



Covariance

- When points on a scatter plot seem to fluctuate around a straight line, there is a linear relationship between *x* and *y*

- A measure of the strength of a linear relationship is the covariance $s_{xy} = \frac{\sum_{i-1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$

- A positive covariance indicates a positive linear relationship between *x* and *y*
  - As *x* increases, *y* increases

- A negative covariance indicates a negative linear relationship between *x* and *y*
  - As *x* increases, *y* decreases

Correlation Coefficient

Correlation measures both the strength and direction of the linear relationship between two variables.

$$Correlation = \frac{Cov\,(x, y)}{\sigma x * \sigma y}$$

- Sample correlation coefficient *r* is always between -1 and +1
    - Values near -1 show strong negative correlation
    - Values near 0 show no correlation
    - Values near +1 show strong positive correlation

Data Presentation

Data presentation is a way to represent the distribution of the variable so that we understand its center, shape, and spread.

There are two ways to present data:

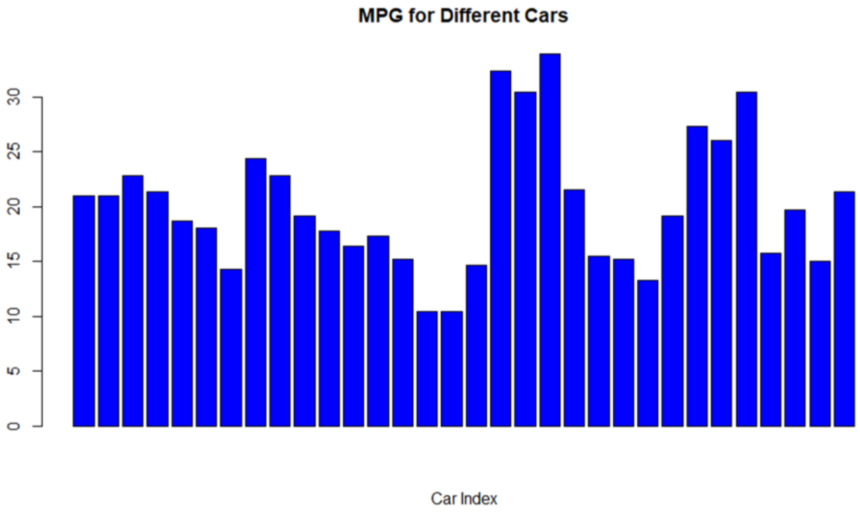
- Graphical presentation: involves some type of chart
- Tabular presentation: involves organizing data into a table

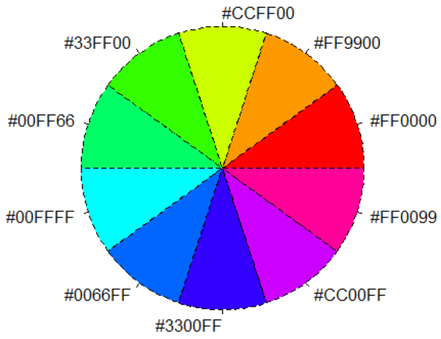
Bar Charts and Pie Charts

Bar chart: A vertical or horizontal rectangle represents the frequency for each category

- Height can be frequency, relative frequency, or percent frequency

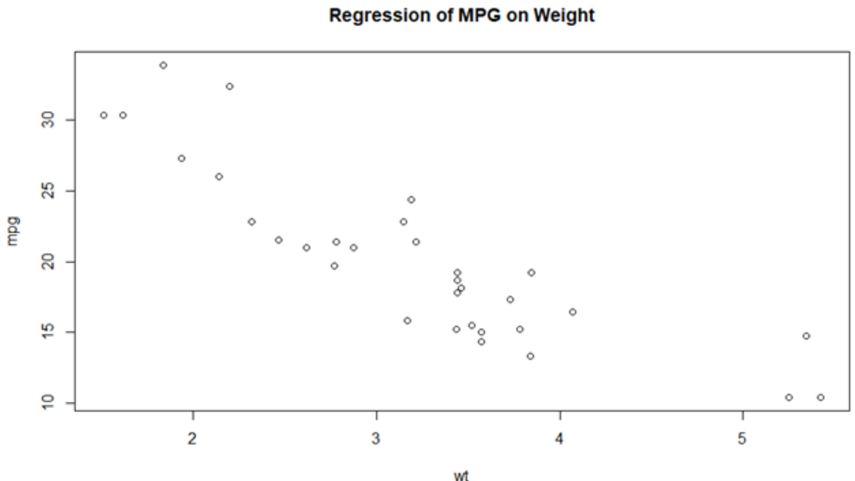
- **Pie chart:** A circle divided into slices where the size of each slice represents its relative frequency or percent frequency





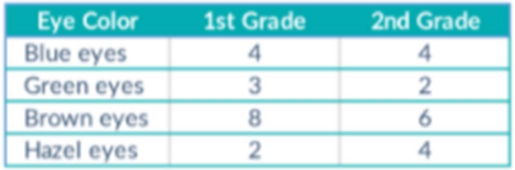
Scatter Plots

- Used to study relationships between two variables
- Place one variable on the x-axis
- Place a second variable on the y-axis
- Place dot on pair coordinates

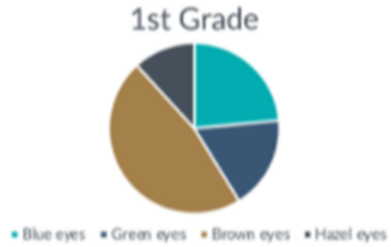


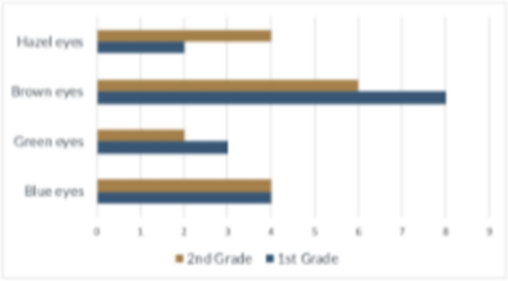
Data Presentation

Categorical Variables

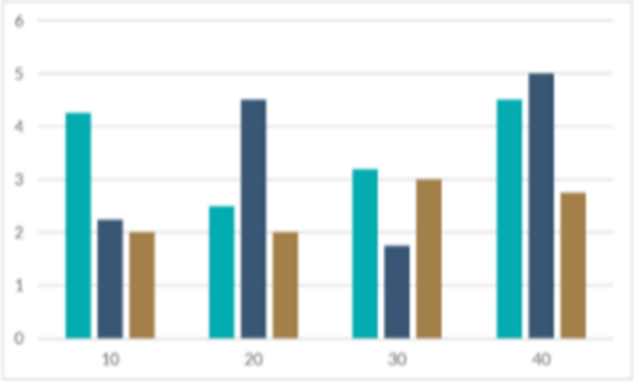


Eye Color	1st Grade	2nd Grade
Blue eyes	4	4
Green eyes	3	2
Brown eyes	8	6
Hazel eyes	2	4

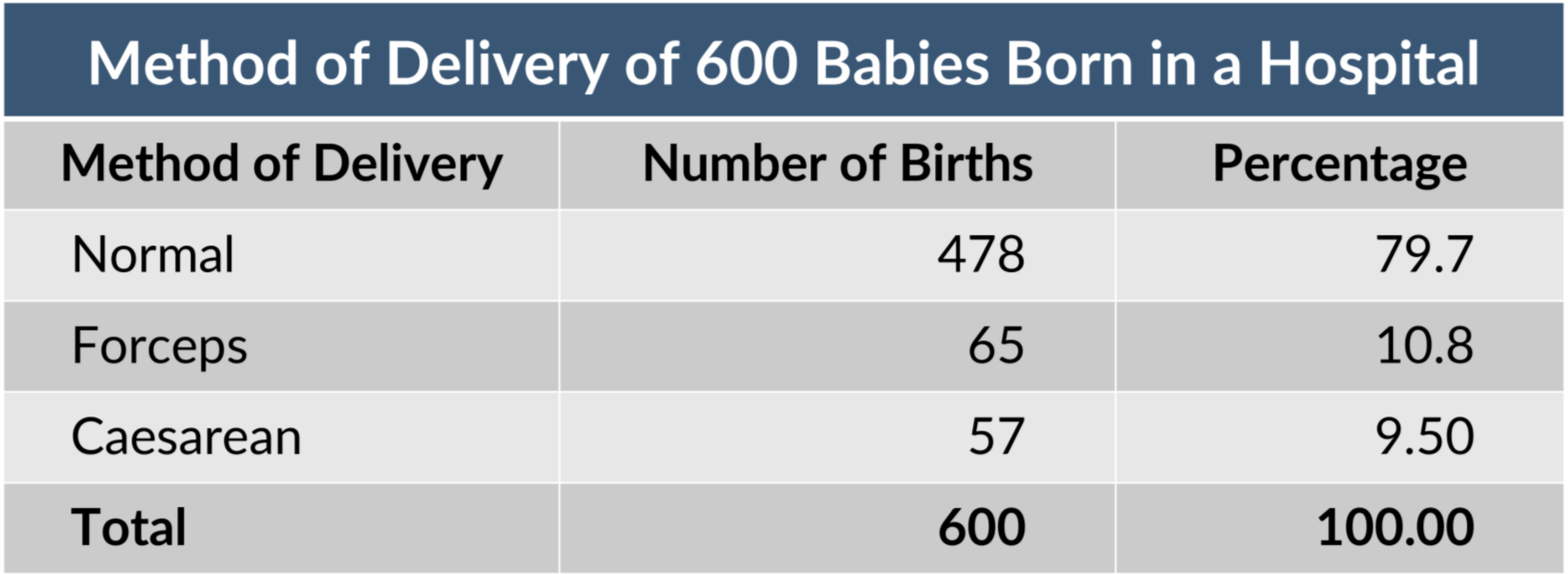




Numerical Variables

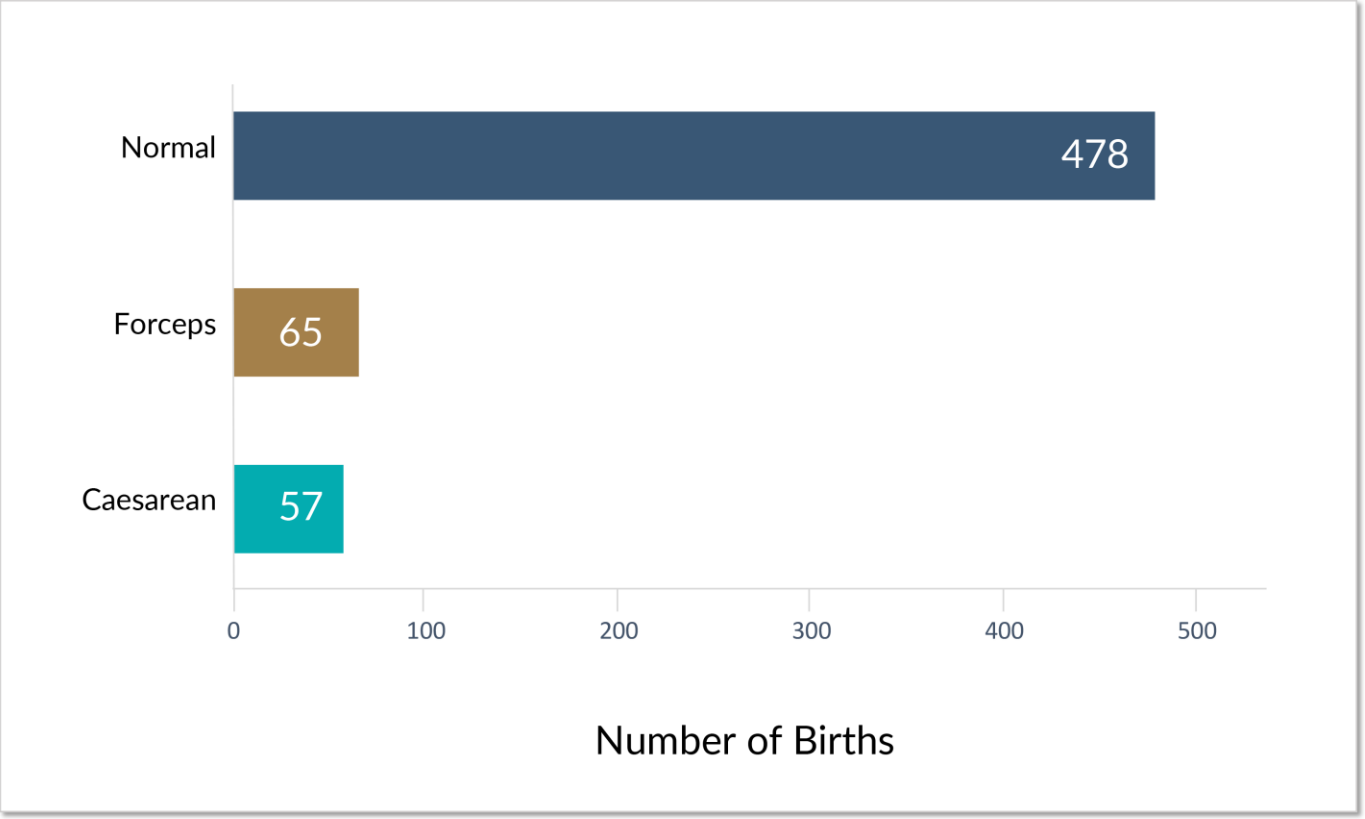


Categorical Data: Table

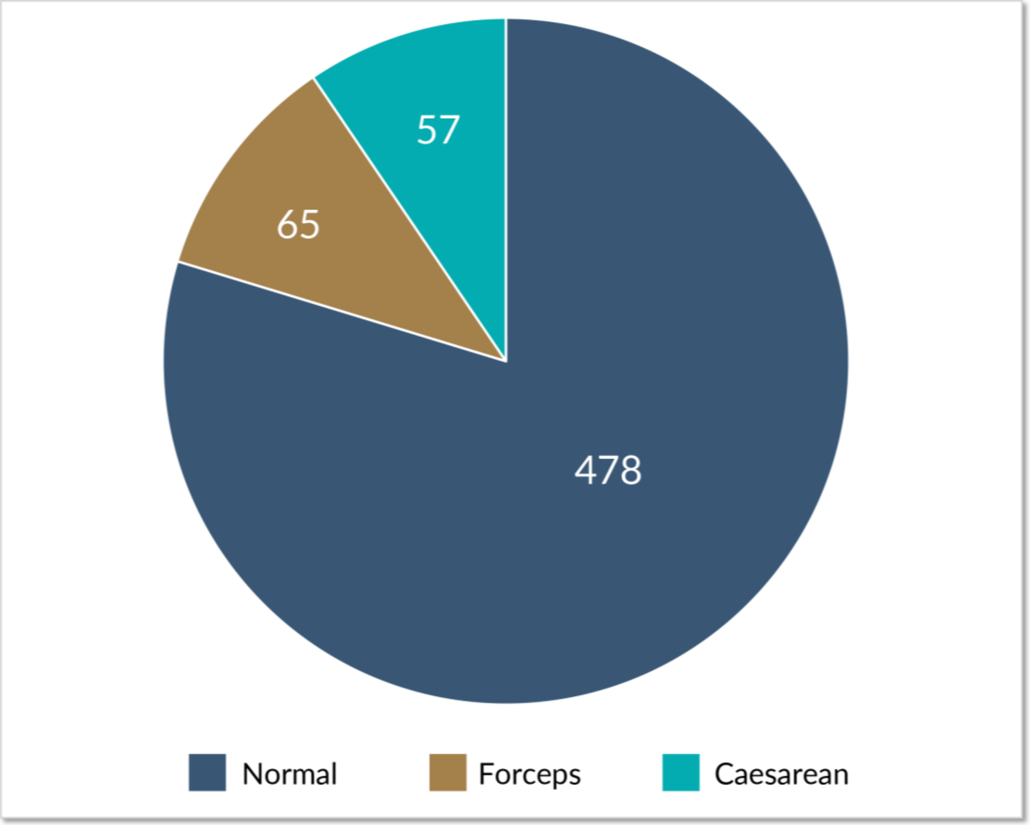


Method of Delivery of 600 Babies Born in a Hospital		
Method of Delivery	**Number of Births**	**Percentage**
Normal	478	79.7
Forceps	65	10.8
Caesarean	57	9.50
Total	**600**	**100.00**

Categorical Data: Bar Chart

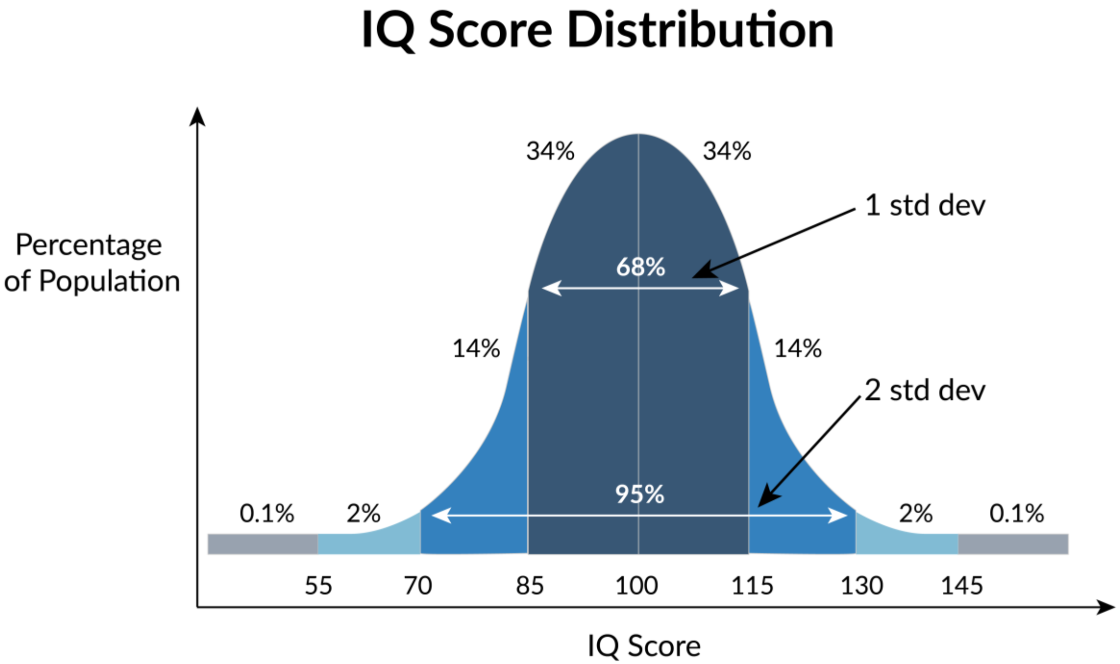


Categorical Data: Pie Chart



Numerical Variables - Distributions

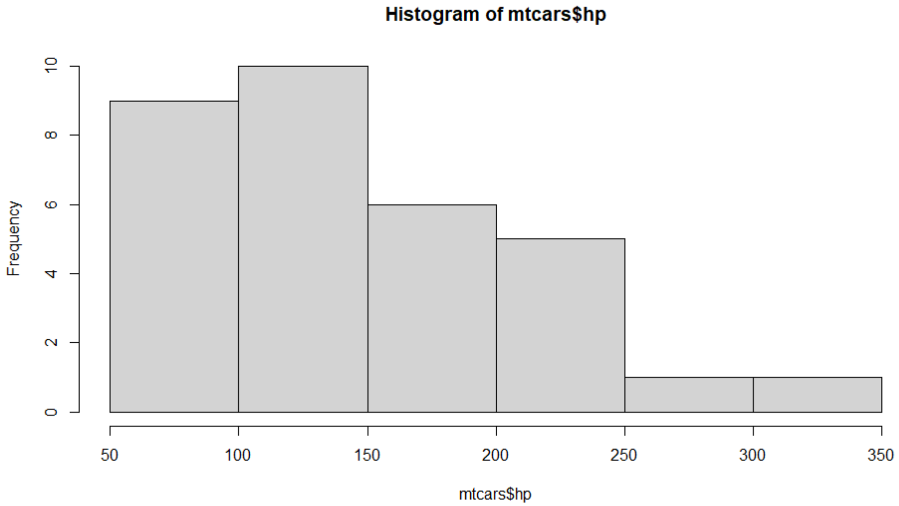
- The graphic below displays the distribution of numerical data.
- The x-axis displays the IQ score, while the y-axis shows the percentage of the population that has that score.
- The distribution shows a bell curve – highest in the middle, symmetric, and tapering on both ends.



Numerical Variables - Histograms

- One of the most useful charts for numerical data is a histogram. A histogram is a bar chart presentation of the frequency table.

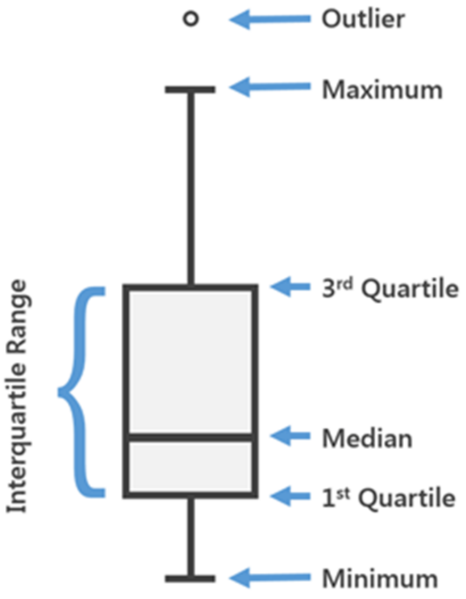
- The data is binned, as we described for the frequency table with numerical data, and the counts for each bin are displayed graphically. In a histogram, the y-axis is always frequency or the number of data points that fall in the given bin.



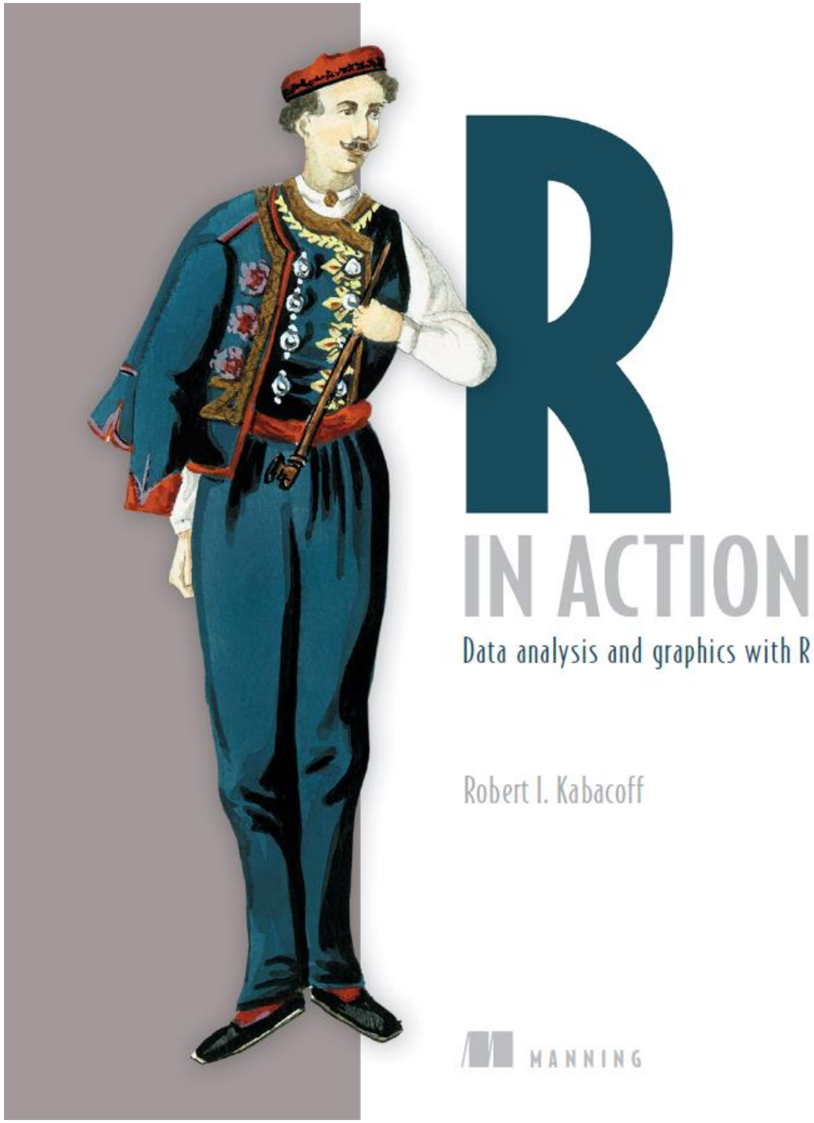
Numerical Variables - Box Plots

Another useful plot for numerical data is the box plot. It displays the five-number summary of the variable: the minimum, maximum, and median are depicted via horizontal lines on the plot.

The box indicates the first, second, and third quartiles. The data points shown as circles are outliers. These terms will be described in greater detail shortly.



R in Action


R
IN ACTION
Data analysis and graphics with R
Robert I. Kabacoff
MANNING

Common Statistical Function

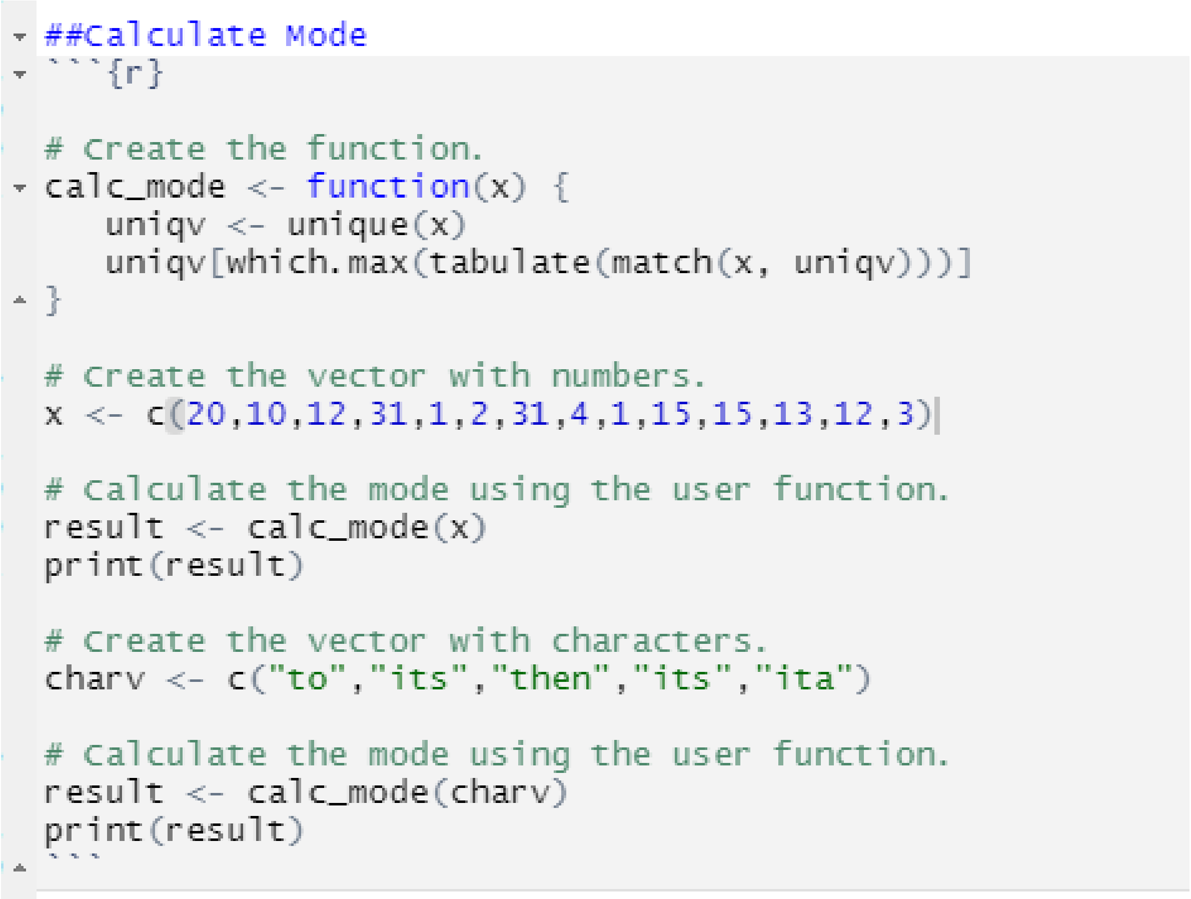
- x<-c(1,2,3,4,5,6,6,5,3,3,33,6)

- mean_Val<-mean(x)
- mean_Val
- median_val<-median(x)
- median_val
- sd_val<-sd(x)
- sd_val

- #install.packages("moments")
- library(moments)
- sk_val<-skewness(x)

- sk_val
- kur_val<-kurtosis(x)
- kur_val

Mode



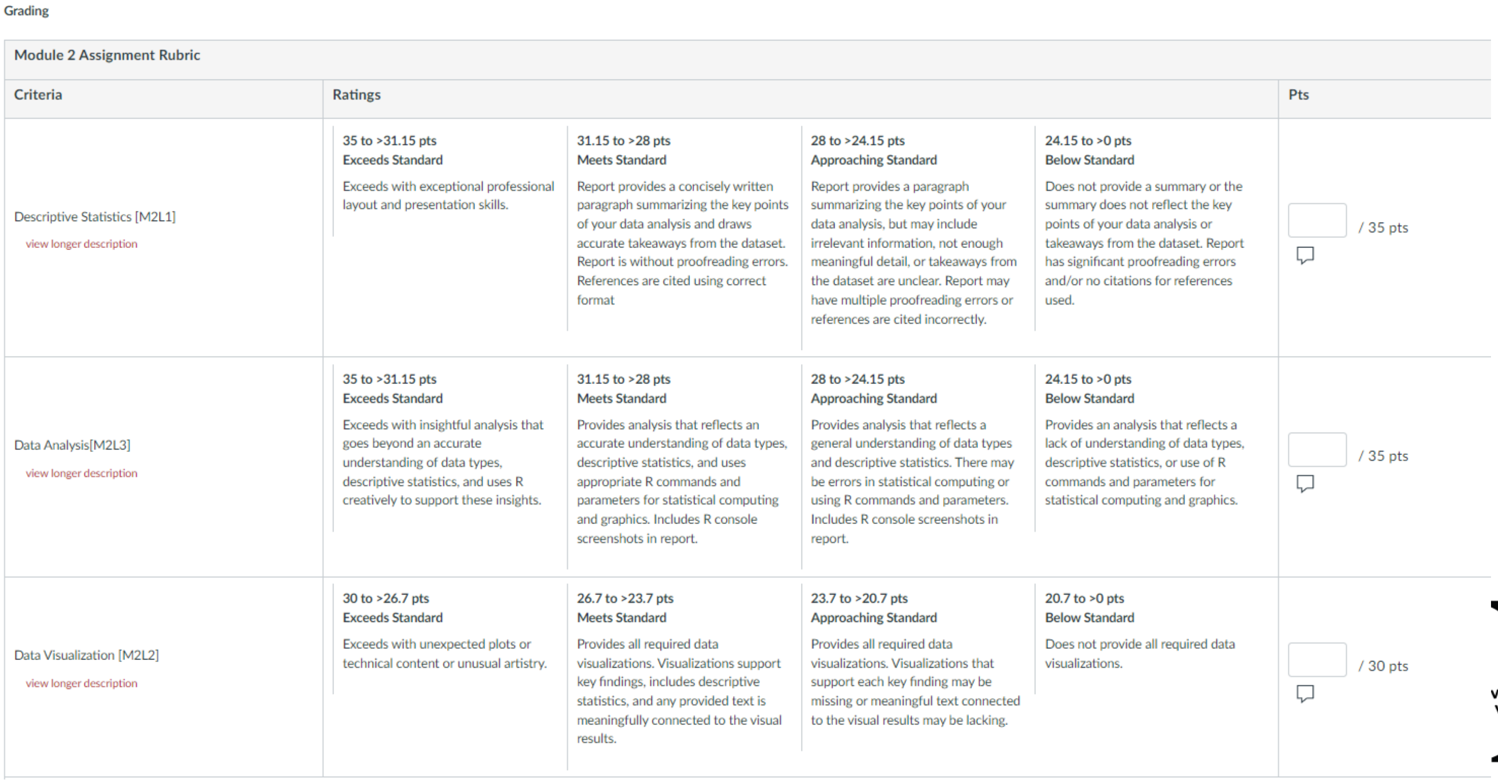
~~~
##Calculate Mode
```{r}

# Create the function.
calc_mode <- function(x) {
    uniqv <- unique(x)
    uniqv[which.max(tabulate(match(x, uniqv)))]
}

# Create the vector with numbers.
x <- c(20,10,12,31,1,2,31,4,1,15,15,13,12,3)

# Calculate the mode using the user function.
result <- calc_mode(x)
print(result)

# Create the vector with characters.
charv <- c("to","its","then","its","ita")

# Calculate the mode using the user function.
result <- calc_mode(charv)
print(result)
```
~~~

Section 3.1 and 3.2

```
# --Section 3.1--

attach(mtcars)
plot(wt, mpg)
abline(lm(mpg ~ wt))
title("Regression of MPG on Weight")
detach(mtcars)

# --Section 3.2--

dose <- c(20, 30, 40, 45, 60)
drugA <- c(16, 20, 27, 40, 60)
drugB <- c(15, 18, 25, 31, 40)
plot(dose, drugA, type = "b")
```

Review of Module Project

- Module 2 Project Instructions
- Executive Summary Report 2



Grading

Module 2 Assignment Rubric

Criteria	Ratings				Pts
Descriptive Statistics [M2L1] view longer description	35 to >31.15 pts **Exceeds Standard** Exceeds with exceptional professional layout and presentation skills.	31.15 to >28 pts **Meets Standard** Report provides a concisely written paragraph summarizing the key points of your data analysis and draws accurate takeaways from the dataset. Report is without proofreading errors. References are cited using correct format	28 to >24.15 pts **Approaching Standard** Report provides a paragraph summarizing the key points of your data analysis, but may include irrelevant information, not enough meaningful detail, or takeaways from the dataset are unclear. Report may have multiple proofreading errors or references are cited incorrectly.	24.15 to >0 pts **Below Standard** Does not provide a summary or the summary does not reflect the key points of your data analysis or takeaways from the dataset. Report has significant proofreading errors and/or no citations for references used.	/ 35 pts
Data Analysis[M2L3] view longer description	35 to >31.15 pts **Exceeds Standard** Exceeds with insightful analysis that goes beyond an accurate understanding of data types, descriptive statistics, and uses R creatively to support these insights.	31.15 to >28 pts **Meets Standard** Provides analysis that reflects an accurate understanding of data types, descriptive statistics, and uses appropriate R commands and parameters for statistical computing and graphics. Includes R console screenshots in report.	28 to >24.15 pts **Approaching Standard** Provides analysis that reflects a general understanding of data types and descriptive statistics. There may be errors in statistical computing or using R commands and parameters. Includes R console screenshots in report.	24.15 to >0 pts **Below Standard** Provides an analysis that reflects a lack of understanding of data types, descriptive statistics, or use of R commands and parameters for statistical computing and graphics.	/ 35 pts
Data Visualization [M2L2] view longer description	30 to >26.7 pts **Exceeds Standard** Exceeds with unexpected plots or technical content or unusual artistry.	26.7 to >23.7 pts **Meets Standard** Provides all required data visualizations. Visualizations support key findings, includes descriptive statistics, and any provided text is meaningfully connected to the visual results.	23.7 to >20.7 pts **Approaching Standard** Provides all required data visualizations. Visualizations that support each key finding may be missing or meaningful text connected to the visual results may be lacking.	20.7 to >0 pts **Below Standard** Does not provide all required data visualizations.	/ 30 pts

At the end of assignment, you can see the Rubric

Summary

- Reviewed module 1 and related topics
- Introduced descriptive statistics
- Explained various types of data visualizations
- Use R to visualize data
- Reviewed Module Project

Q &A