Stat 201: Introduction to Statistics

Standard 7: Numerical Summaries – Percentiles

Chapter Two

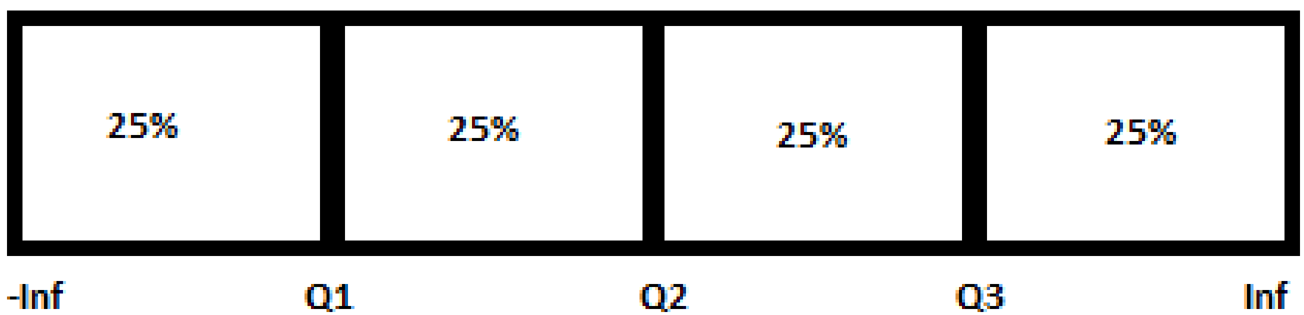
Summaries

Percentiles

- How many of you have heard this term before?
  - Testing
  - Medical terminology
  - Etc
- **Percentiles -** the pth percentile is a value such that p percent of the observations fall below or at that value.

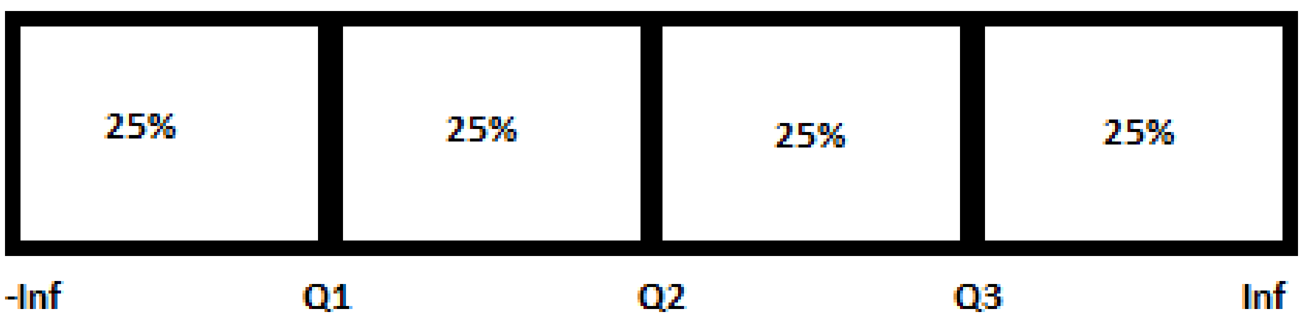
Percentiles: Important Ones

- We call these quartiles because they split the data into quarters
  - Q1: the observation at the 25th percentile
  - Q2: the observation at the 50th percentile
    - This is the same as the median
  - Q3: the observation at the 75th percentile



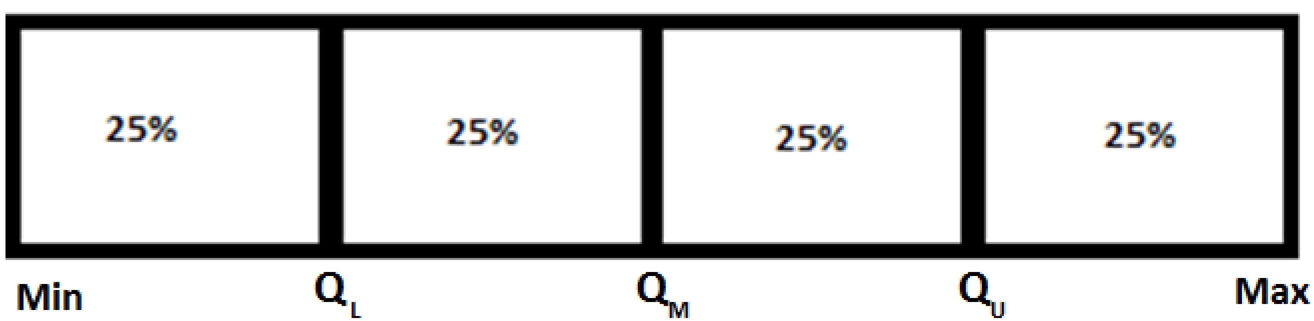
Percentiles: Important Ones

- **IQR**=Q3-Q1: another measure of spread used in place of standard deviation w/ skewed data
  - IQR gives the range of the middle 50% of the data



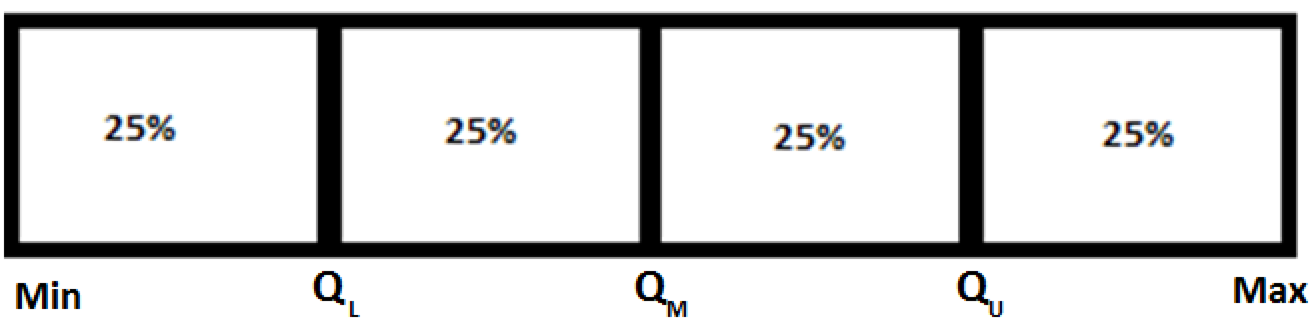
Five Number Summary: Important Percentiles

- We call these quartiles because they split the data into quarters
  - Q_L: the observation at the 25th percentile
  - Q_M: the observation at the 50th percentile
    - This is the same as the median
  - Q_U: the observation at the 75th percentile
- Min: the smallest observation – the 0th percentile
- Max: the largest observation – the 100th percentile



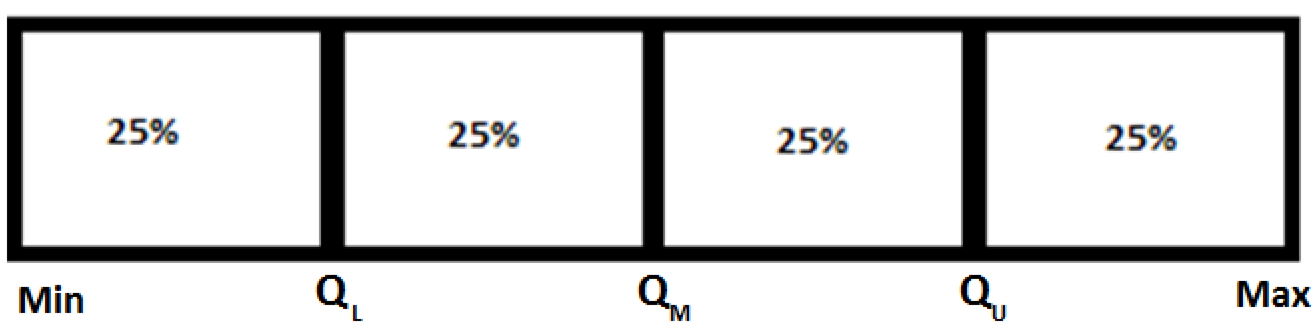
Five Number Summary: Interquartile Range

- **IQR**=Q_U-Q_L: another measure of spread used in place of standard deviation w/ skewed data
  - IQR gives the range of the middle 50% of the data



Five Number Summary: Finding Outliers with Quartiles

- Lower Fence= Q_L – (1.5)*IQR
  = 1.5 – (1.5)*5 = -6
- Upper Fence= Q_U + (1.5)*IQR
  = 6.5 + (1.5)*5 = 14
- **We consider any observation with a value outside of the interval (Lower Fence, Upper Fence) an outlier**



Walkthrough

Percentiles

- How many of you have heard this term before?
  - Testing
  - Medical terminology
  - Etc
- **Percentiles -** the pth percentile is a value such that p percent of the observations fall below or at that value.

Five Number Summary: Where to Find Them

- The five number summary, of n items, that we use to draw a box plot includes the following:

Name	Position in Ascending Order
Minimum	1^{st}
Q_1	$.25*(n+1)^{th}$
Q_M (This is the median)	$.5*(n+1)^{th}$
Q_3	$.75*(n+1)^{th}$
Maximum	n^{th}

Example: The Lower (1st) Quartile

Is the position value a whole number	The Quartile
Yes	The number in that position
No	The weighted average of the numbers in the above and below positions

- X = {0,1,2,3,4,5,6,7,8)
- Position of Q_1 = .25*(n+1) = .25*(9+1)

  = 2.5th position (the remainder is .5)
- Q_1= (.5)*(# In the 3rd pos.) +(1-.5)*(# in the 2nd pos.)

  = .5*2 + .5*1 = 1 + .5 = 1.5

Example: The Middle (2^{nd}) Quartile

Is the position value a whole number	The Quartile
Yes	The number in that position
No	The average of the numbers in the above and below positions

- X = {0,1,2,3,4,5,6,7,8)
- Position of the Median = .5*(n+1) = .5*(9+1)
  = 5^{th} position
- Q_M= 4

Example: The Upper (3^{rd}) Quartile

Is the position value a whole number	The Quartile
Yes	The number in that position
No	The average of the numbers in the above and below positions

- X = {0,1,2,3,4,5,6,7,8)
- Position of Q_3 = .75*(n+1) = .75*(9+1)

  = 7.5^{th} position (.5 is the remainder)
- Q_3 = (.5)*(# In the 8^{th} pos.) +(1-.5)*(# in the 7^{th} pos.)

  = .5*7 + .5*6 = 1 + 1.5 = 6.5

Example: Interquartile Range

X = {0,1,2,3,4,5,6,7,8)

- Q_1 = (1+2)/2 = 1.5
- Q_M = 4
- Q_3 = (6+7)/2 = 6.5

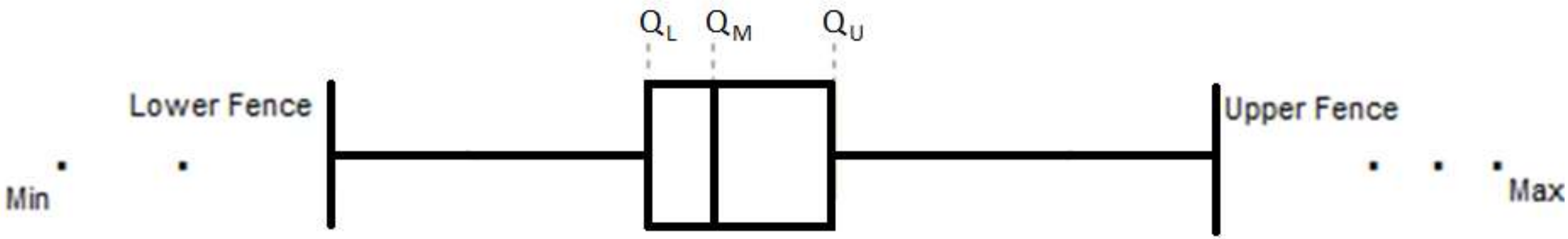
- IQR = $Q_3 - Q_1$ = 6.5 - 1.5 = 5
  - 50% of the data lies between 1.5 and 6.5
  - 50% of the data lies on a range of size 5

Example: Using Quartiles to find Outliers

X = {0,1,2,3,4,5,6,7,8)

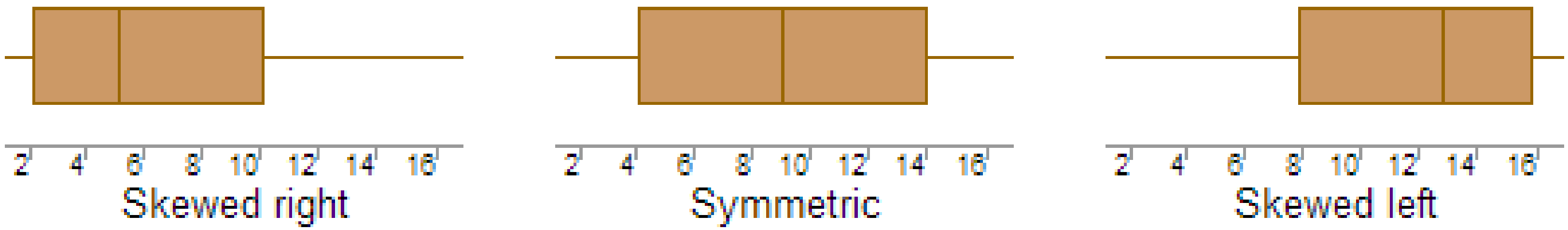
- Q_1 = (1+2)/2 = 1.5
- Q_3 = (6+7)/2 = 6.5
- IQR = $Q_3 - Q_1$ = 6.5 - 1.5 = 5
- Lower Fence= Q_1 – (1.5)*IQR
  = 1.5 – (1.5)*5 = -6
- Upper Fence= Q_3 + (1.5)*IQR
  = 6.5 + (1.5)*5 = 14
- **In this case anything smaller than -6 or greater than 14 would be an outlier**

Box Plots:
The Graph of a Five Number Summary

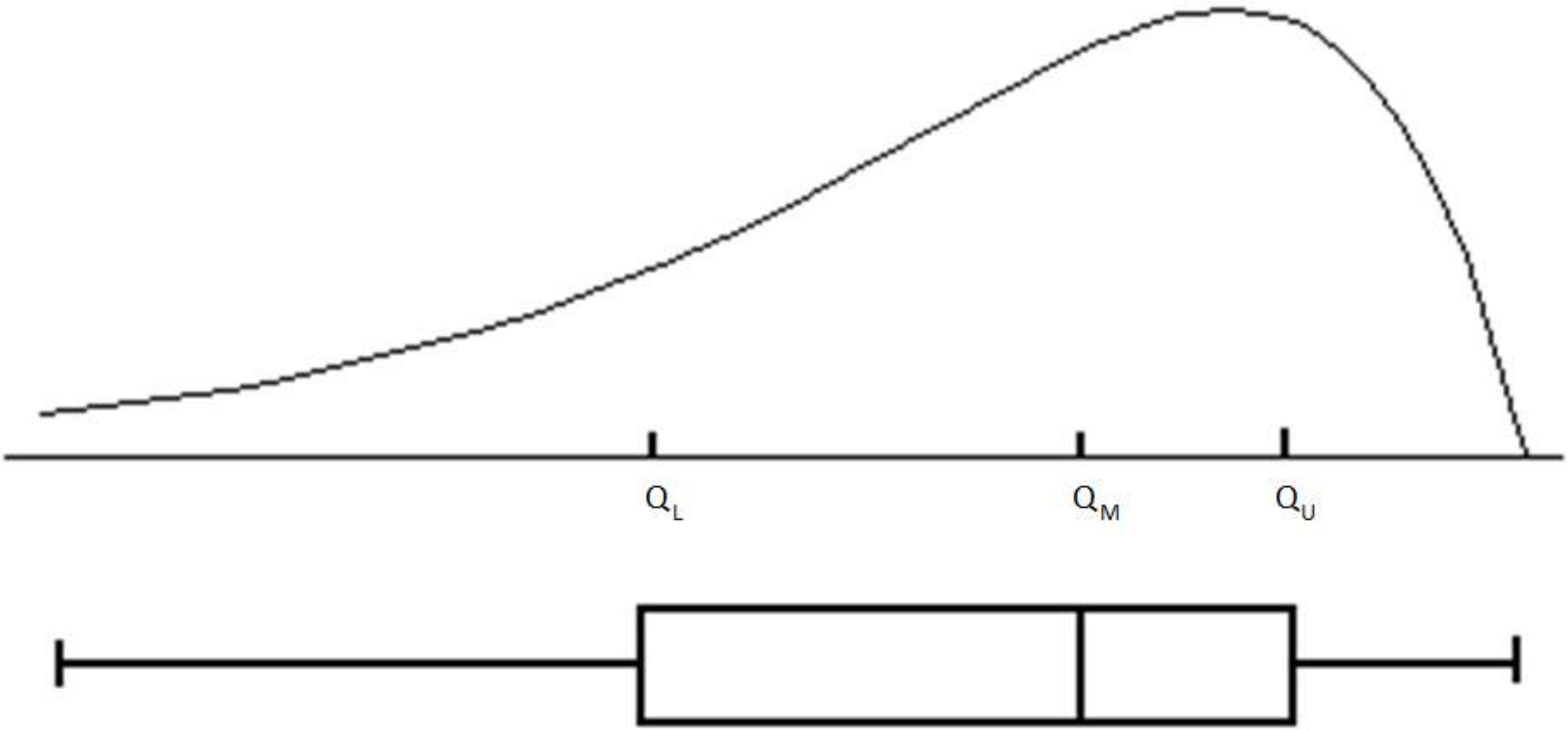


- The box plot utilizes the five number summary
  - The box is created using quartiles
  - The whiskers are created using the fences
  - The points are the outlying points –if there are any

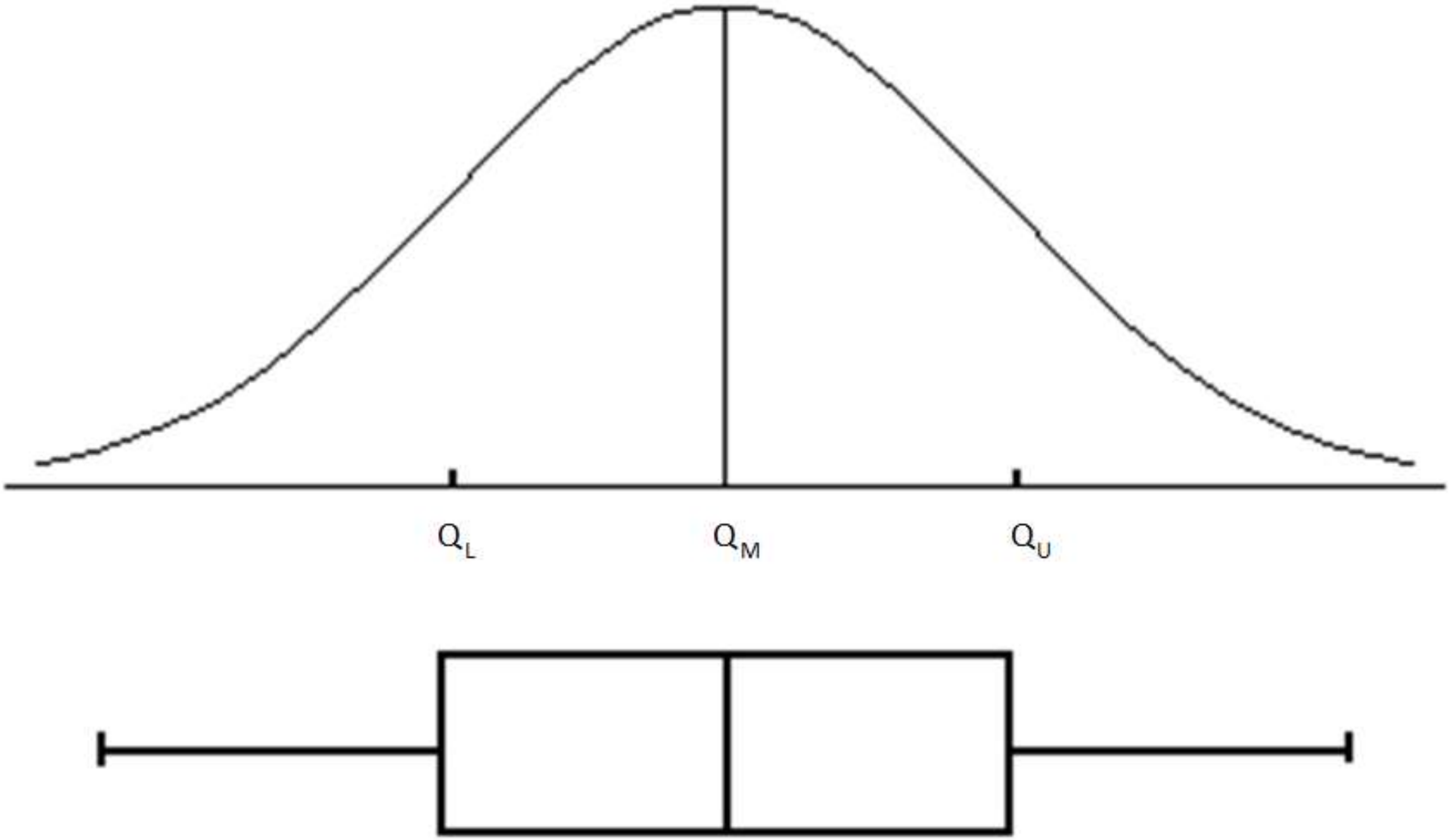
Skewness in Boxplots



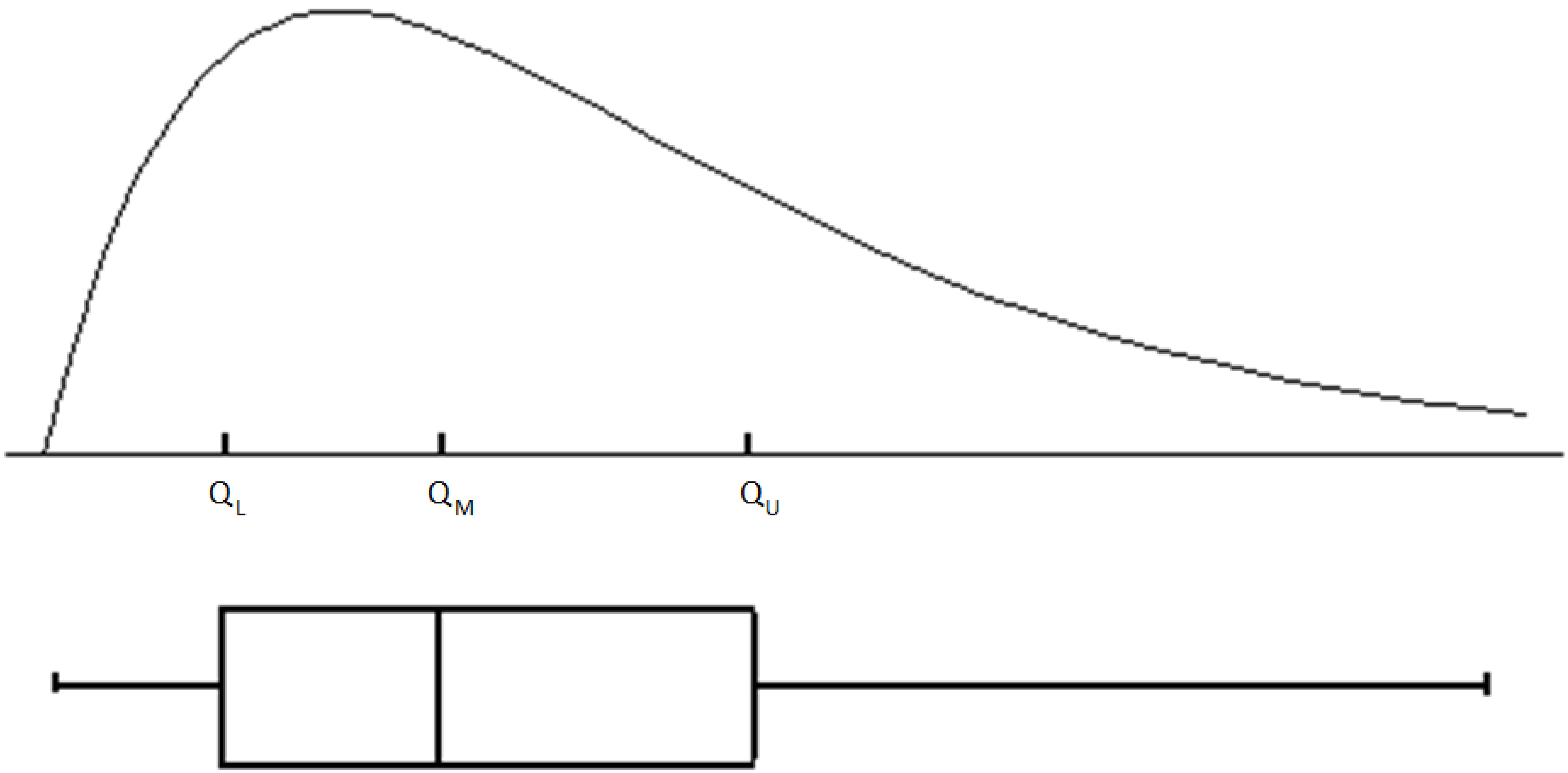
Left Skewed w/ Boxplots

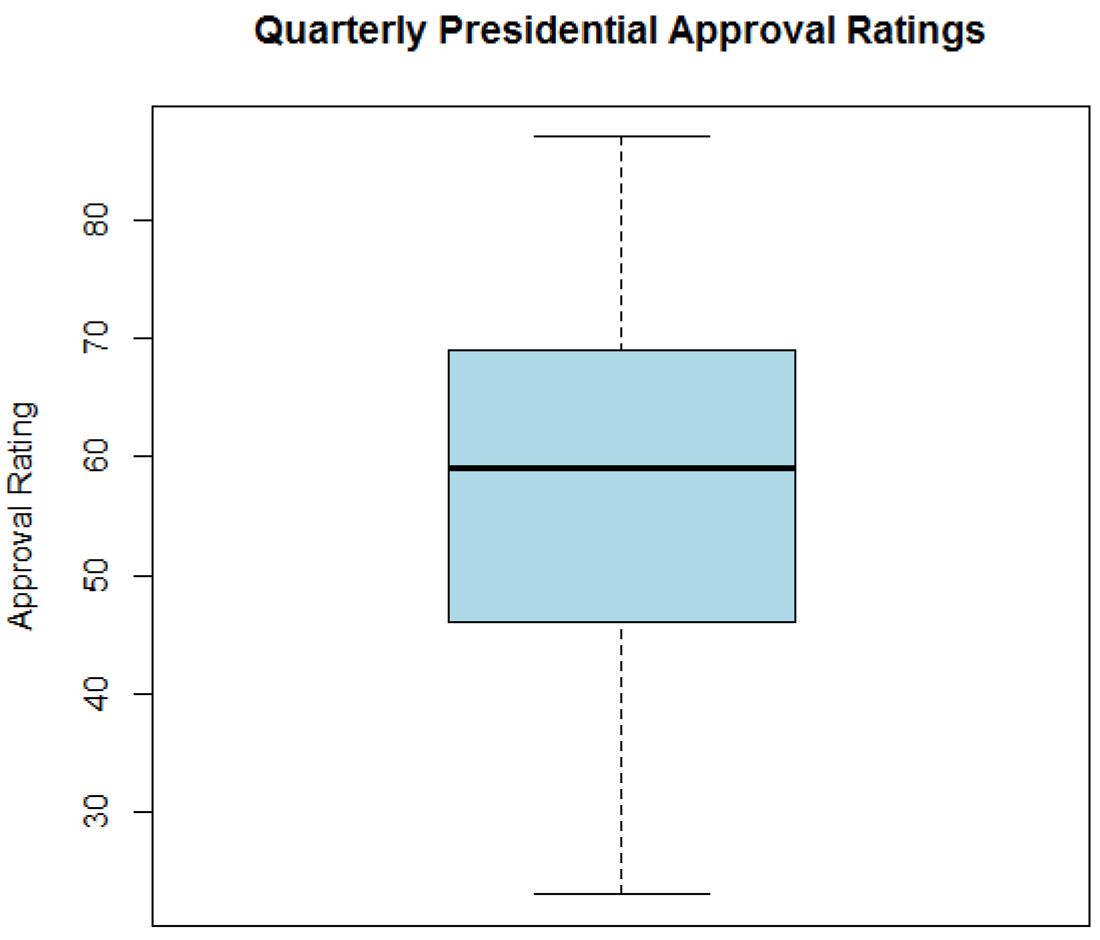


Bell Shaped w/ Boxplots



Right Skewed w/ Boxplots



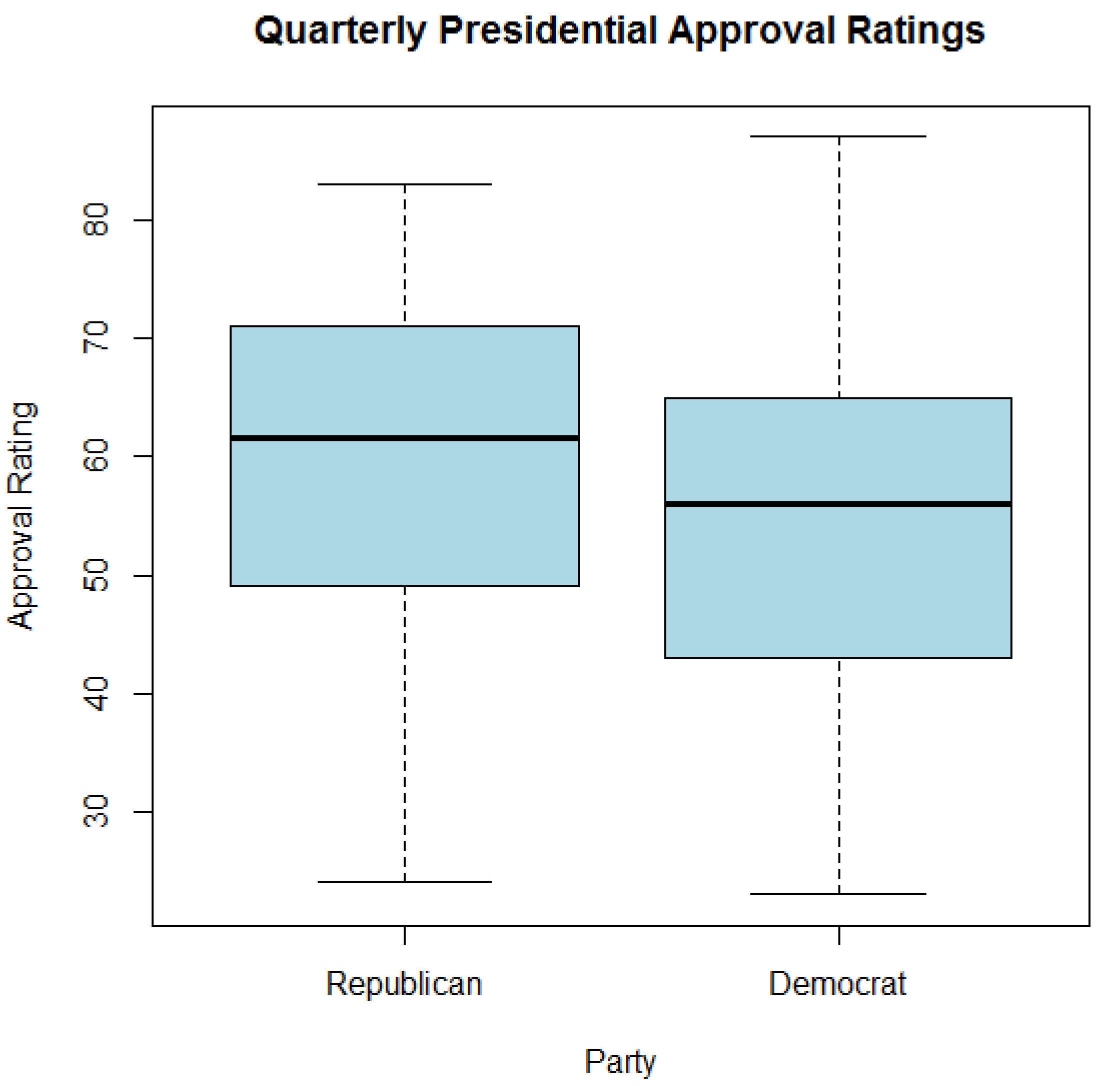


Quarterly Presidential Approval Ratings
Approval Rating
80
70
60
50
40
30

Data: Graphical Summary

- **StatCrunch Command:**

Graph→Boxplot→ Select the variable(s)→Compute



Quarterly Presidential Approval Ratings
Approval Rating
30
40
50
60
70
80
Republican
Democrat
Party

Data: Graphical Summary

- **StatCrunch Command**

Graph→Bar Plot→w/data→ Select the variable you'd like on the x-axis→Group by the variable you would like the bars to be split by→Compute