MATH 2311

Section 1.5

Graphs and Describing Distributions

- Lets start with an example:
- Height measurements for a group of people were taken. The
- results are recorded below (in inches): 66, 68, 63, 71, 68, 69, 65,
- 70, 73, 67, 62, 59, 63, 68, 71, 63, 63, 60, 64, 66, 58

- We will organize this data using different graphs:

Ice Cream Pref	Vanilla	Chocolate	Strawberry	Coffee
Frequency	12	18	16	8

To copy:
66,68,63,71,68,69,65,70,73,67,62,59,63,68,71,63,
63,60,64,66,58

Items in bold are graphs that refer to Quantitative data

Graphs in R-Studio

Histograms:

hist(x)

Boxplots:

boxplot(x)

Dot Chart:

dotchart(x)

Stem and Leaf:

stem(x)

Pie Chart:

pie(x)

Bar Graph:

barplot(x)

Graphs using the TI-83/84

You can only plot Histograms, Box Plots, or Box and Whisker Plots using the TI-83/84 Calculator.

Press: 2^{nd} StatPlot [Y =]

On Plot 1, Press ENTER

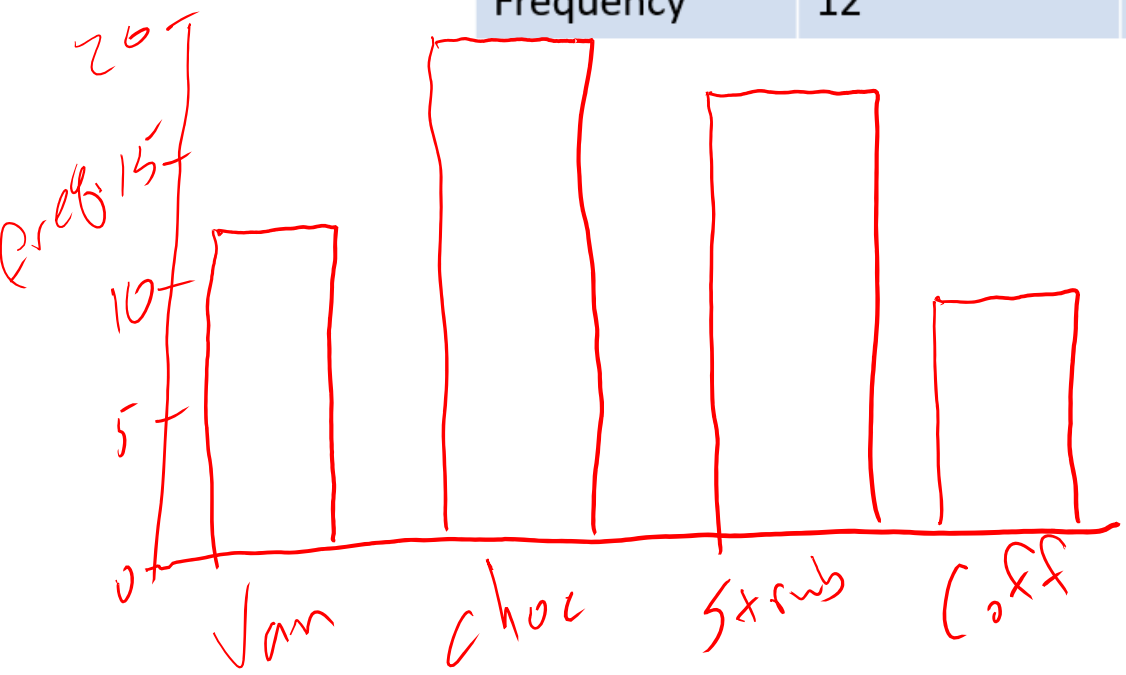
Select On, Select Graph Type, Select appropriate lists or numbers

To view plot, press GRAPH

(Be certain to shut off StatPlot1 to graph normal functions without difficulties.)

A **bar graph** is created by listing the categorical data along the x-axis (horizontal) and the frequencies along the y-axis (vertical). Bars are drawn above each data value.

Ice Cream Pref	Vanilla	Chocolate	Strawberry	Coffee
Frequency	12	18	16	8

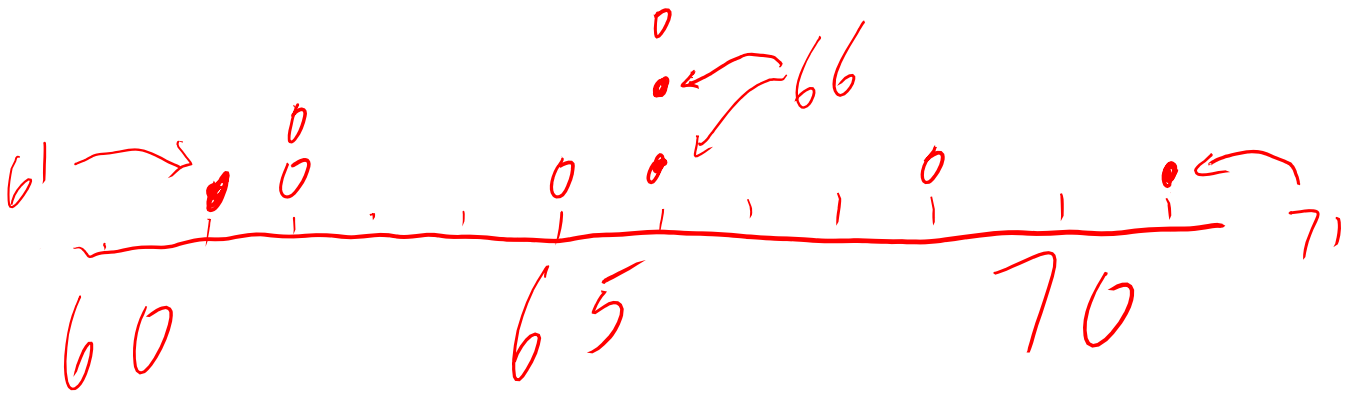


Notes on Bar Graphs: Categorical Data, heights of the individual bars represent frequency of that category, bars do not touch (there is a gap between bars)

A **dot plot** is made simply by putting dots above the values listed on a number line. Quantitative

● : Males

○ : Females



A **stem and leaf plot**, the data is arranged by values. The digits in the largest place are referred to as the stem and the digits in the smallest place are referred to as the leaf (leaves). The leaves are displayed to the right of the stem. A **split stemplot** divides up the stems into equal groups. **Back-to-back stempots** can be used when comparing two sets of data.

Quantitative data, Values: you do not lose any of the specific data points when viewing it

```
> stem(x)

  The decimal point is 1 digit(s) to the right of the |

  5 | 89
  6 | 0233334
  6 | 56678889
  7 | 0113
```

(Split stem and leaf plot)

58, 59

60, 62, 63, 63, 63, 63, 64

65, 66, 66, 67, 68, 68, 68, 69

70, 71, 71, 73

or

5 | 8 = 58

Stem and leaf (Not split)

5	8 9
6	0 2 3 3 3 4 5 6 6 7 8 8 8 9
7	0 1 1 3

Back to Back Stem

9	5	8
4 3	6	0 2 3 3 3
	7	

5 | 8 = Male, 58

9 | 5 = Female, 59

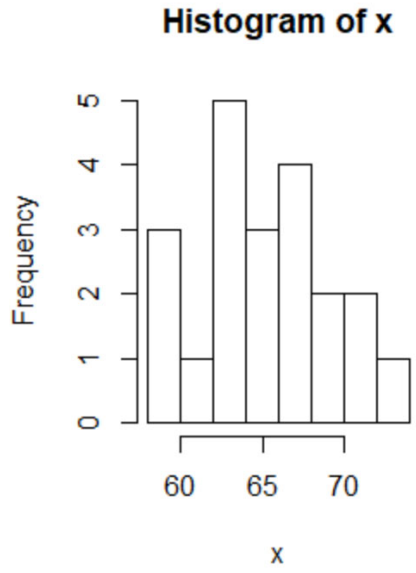
Quantitative Data (Lose individual data points)

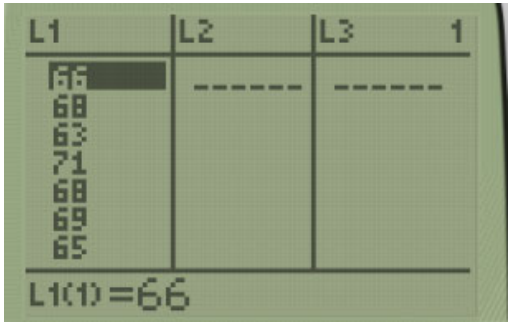
Histograms are created by first dividing the data into classes, or bins, of equal width. Next, count the number of observations in each class. The horizontal axis will represent the variable values and the vertical axis will represent your frequency or your relative frequency.

```
> hist(x)
Error in plot.new() : figure margins too large
> hist(x)
```

Error: Readjust right window, repeat

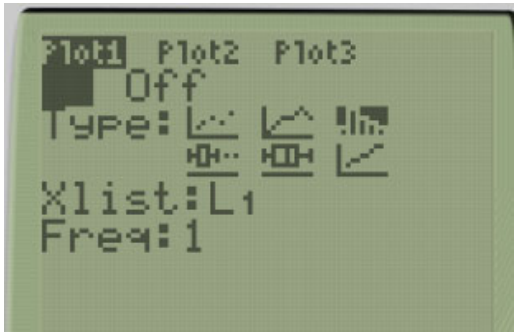
Notes on histograms: height of bars represent frequency of the bin or class. You cannot determine individual date points from a histogram

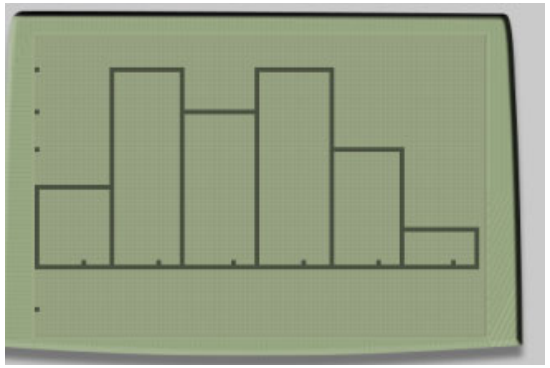



L1 L2 L3 1
66
68
63
71
68
69
65

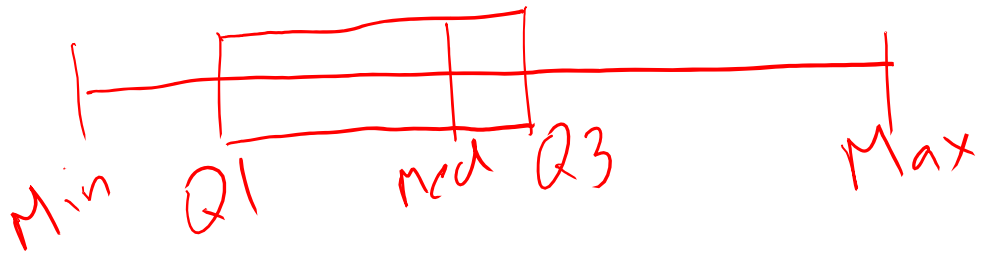
L1(1)=66

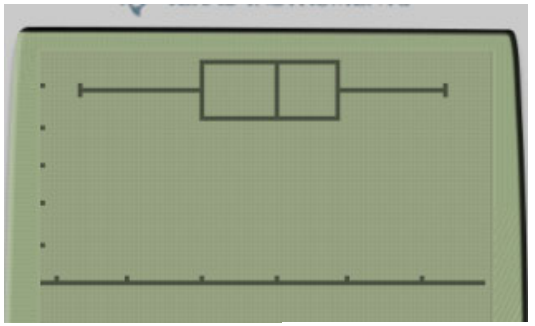

ZOOM MEMORY
3:Zoom Out
4:ZDecimal
5:ZSquare
6:ZStandard
7:ZTrig
8:ZInteger
9:ZoomStat


Plot1 Plot2 Plot3
Off
Type:
Xlist:L1
Freq:1



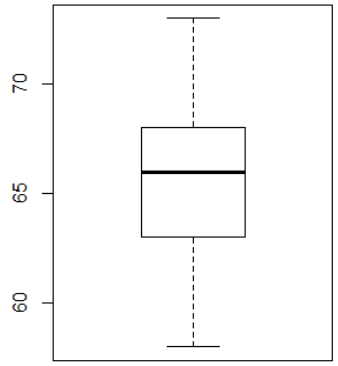
Boxplots not only help identify features about our data quickly (such as spread and location of center) but can be very helpful when comparing data sets. Quantitative Data





Notes on Box-plots: Vertical line at min, Q1, Med, Q3, Max. These should only be symmetric if your data set is symmetric (draw a number line first). The Box represents your IQR, or the middle 50% of data. (If outliers are present, use a box and whisker plot, explained later)

```
> boxplot(x)
```



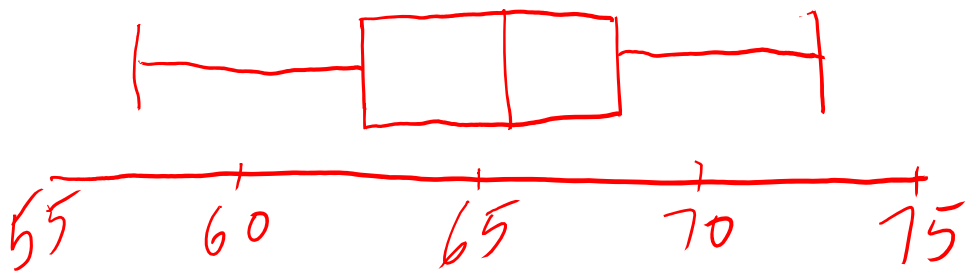
How to make a box plot:

1. Order the values in the data set in ascending order (least to greatest).
2. Find and label the median.
3. Of the lower half (less than the median—do not include), find and label Q1.
4. Of the upper half (greater than the median—do not include), find and label Q3.
5. Label the minimum and maximum.
6. Draw and label the scale on an axis.
7. Plot the five number summary.
8. Sketch a box starting at Q1 to Q3.
9. Sketch a segment within the box to represent the median.
10. Connect the min and max to the box with line segments.

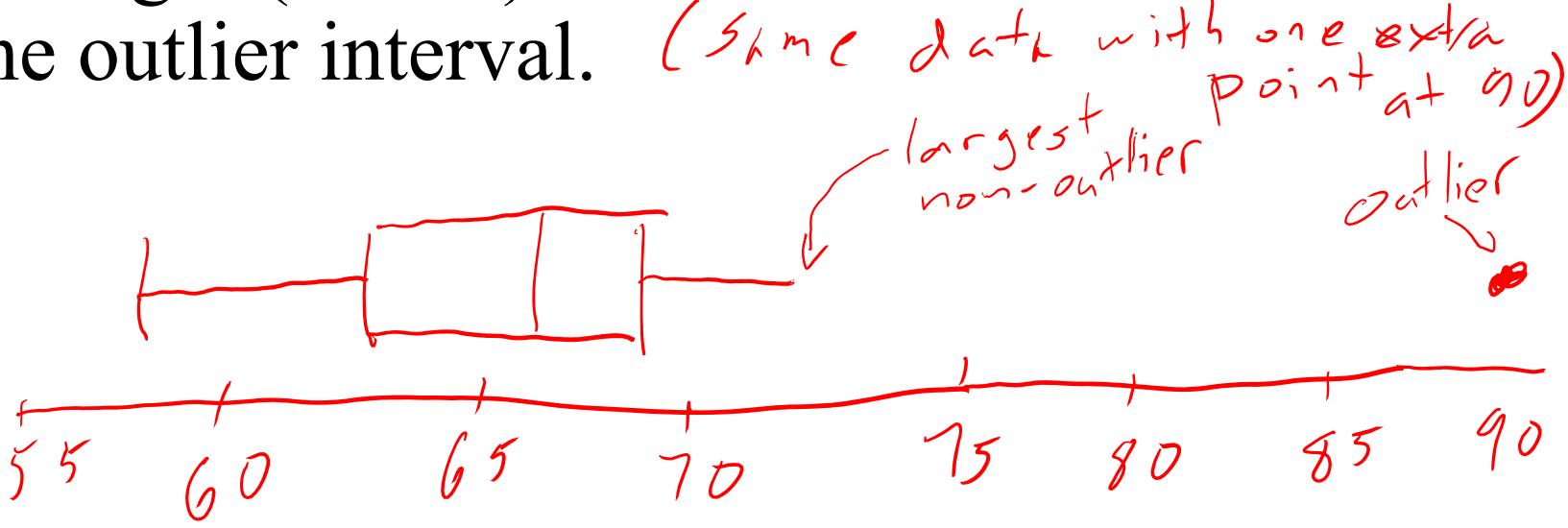
Five Number Summary

```
fivenum(x)
[1] 58 63 66 68 73
```

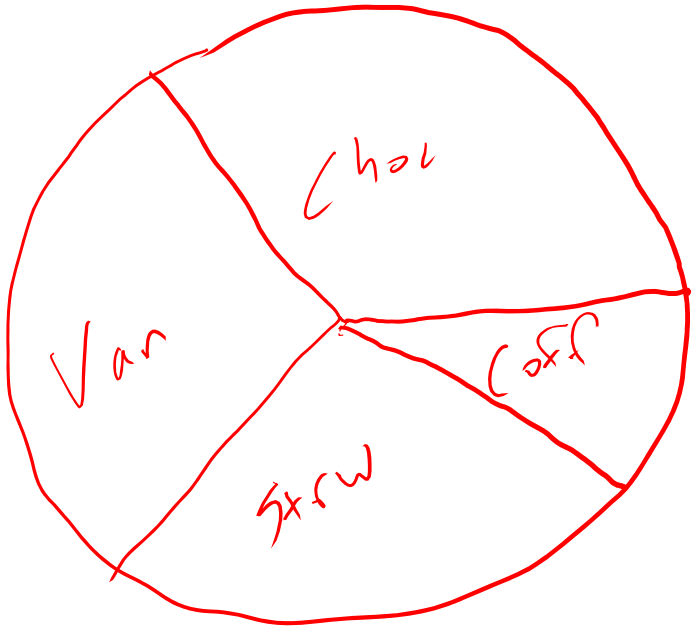
min Q1 med Q3 Max



Note: If data contains outliers, a **box and whiskers plot** can be used instead to display the data. In a box and whiskers plot, the outliers are displayed with dots above the value and the segments begin (or end) at the next data value within the outlier interval.



A **pie chart** is a circular chart, divided into sectors, indicating the proportion of each data value compared to the entire set of values. Pie charts are good for categorical data.

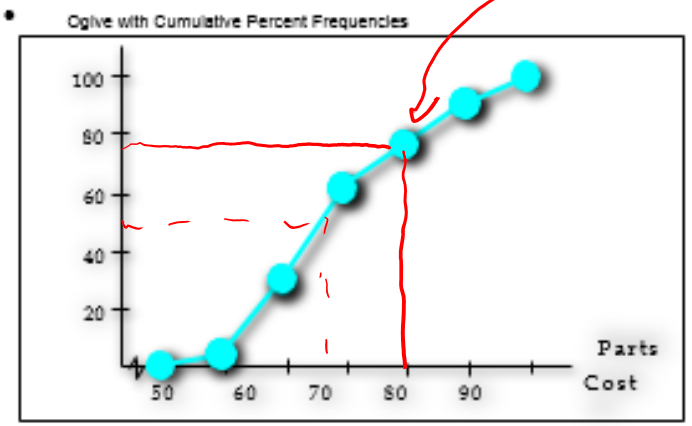


Every category is represented by its own sector. The area of each sector represents the relative frequency of that category.

A **cumulative frequency plot** of the percentages (also called an **ogive)** can be used to view the total number of events that occurred up to a certain value.

Example: Here is an ogive for Hudson Auto Repair's cost of parts sold:

Example: Hudson Auto Repair



78% of repairs are less than $80

Notes of cumulative frequency or ogive: Graphs are always increasing due to their cumulative nature. First data point is always a frequency of 0 and the last data point always a frequency of 100% (or max-value)

Where is the median of this data?

50% marker (since that will represent the cost that 50% of repairs fall beneath)

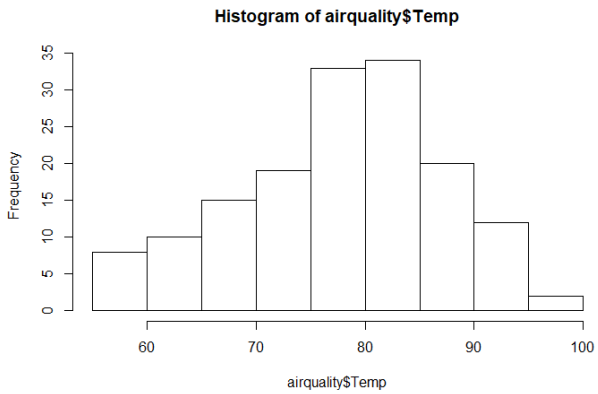
Approx $72

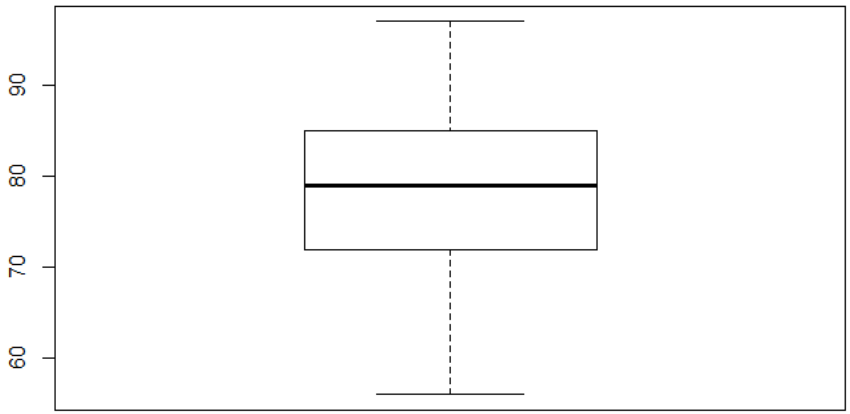
Try this (using R-Studio data packages):

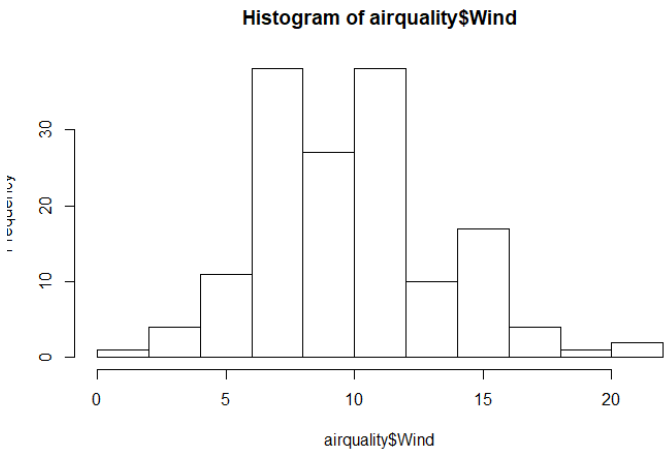
There is a data package (pre-installed in Rstudio) called airquality. This tells daily measurements of air quality readings (ozone, solar level, wind speed, temperature) in New York from May to September of 1973. Command(filename$column)

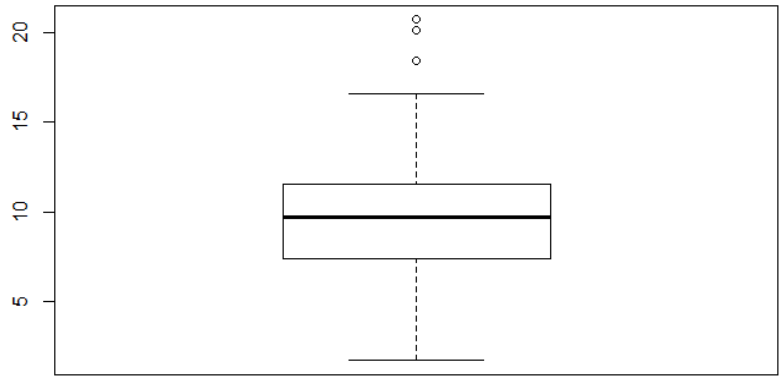
Create a histogram and box plot of wind speed and temperature (these variables are named “wind” and “temp”.

```
> hist(airquality$Temp)
> boxplot(airquality$Temp)
> hist(airquality$Wind)
> boxplot(airquality$Wind)
> mean(airquality$Wind)
[1] 9.957516
> fivenum(airquality$Wind)
[1]  1.7  7.4  9.7 11.5 20.7
```


Histogram of airquality$Temp
Frequency
0
5
10
15
20
25
30
35
60
70
80
90
100
airquality$Temp

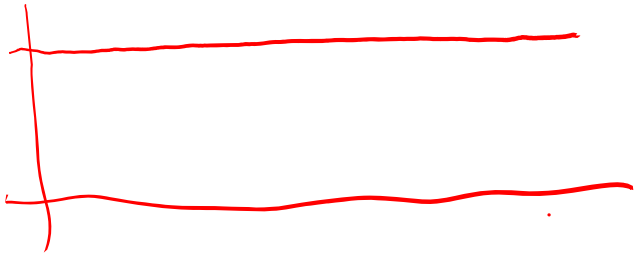

60
70
80
90


Histogram of airquality$Wind
0
10
20
30
0
5
10
15
20
airquality$Wind

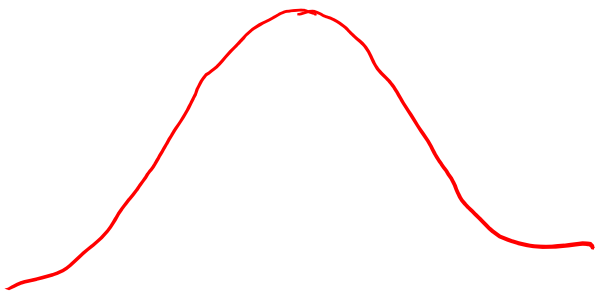

5
10
15
20

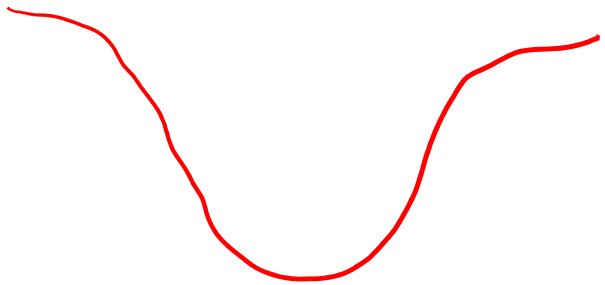
Patterns and shapes:

Uniform graphs All frequencies are equal

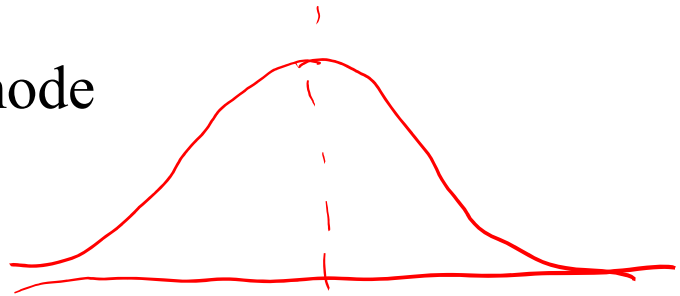


Symmetric graphs: has line symmetry at the center mean=median

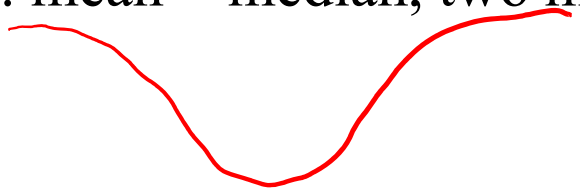




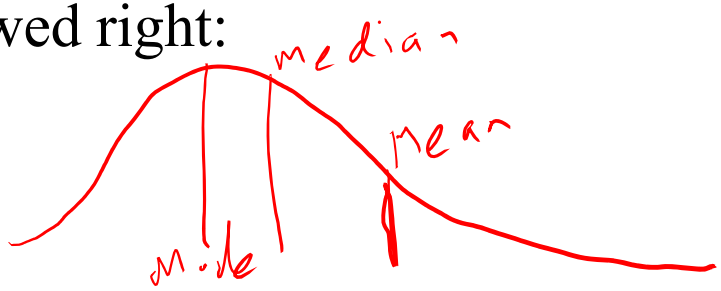
Bell Shaped: mean = median = mode



Bimodal: mean = median, two modes (at either extreme)



Skewed right: mode < median < mean



Skewed left mean < median < mode