Group Work - Chapter 2

1 The high temperatures in Saint Paul for each day in December are below. The data can also be found on D2L as "max_temps_dec17.csv". Use classes -9 to 0, 1 to 10, 11 to 20 etc.

51, 46, 49, 57, 27, 19, 22, 31, 27, 29, 35, 24, 29, 24, 29, 30, 28, 42, 38, 23, 24, 25, 22, 22, 3, 0, 4, 11, 7, −6, −8

(a) Build a frequency table of the data. Do the data appear to have a normal distribution? Are there outliers? Add relative frequency and cumulative frequency tables.

The distribution does seem to be approximately normal by our current definition. There do not appear to be any outliers.

	Freq	Rel	Cum
(-10,0]	3	0.0968	3
(0,10]	3	0.0968	6
(10,20]	2	0.0645	8
(20,30]	15	0.4839	23
(30,40]	3	0.0968	26
(40,50]	3	0.0968	29
(50,60]	2	0.0645	31

(b) Sketch a histogram of the data. Does the histogram show a normal distribution? If not, which characteristic of normal distributions does it violate?

The histogram appears approximately normal.

Max Temps for December 2017
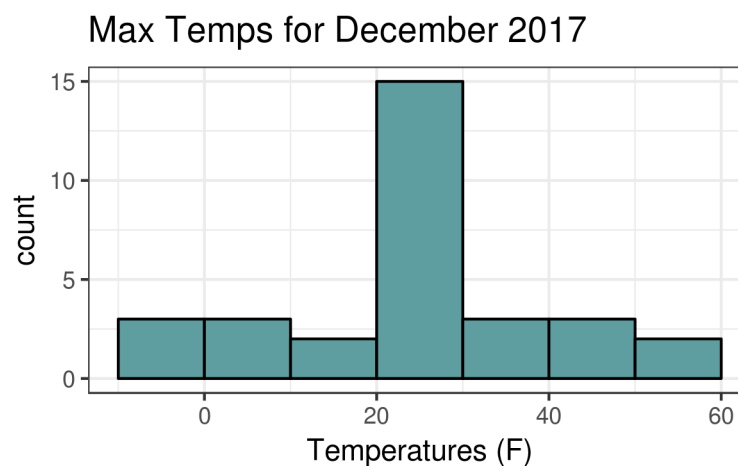
2 The maximum monthly fine particulate matter pollution in Minnesota from 2007 to 2009 is below. The data can be found on D2L as "max_air_pol.csv". Use classes 10-15, 15-20, 20-25, etc.

25.82, 28.87, 36.18, 21.22, 25.94, 25.10, 24.49, 36.94, 23.20, 25.72, 27.94, 43.34

31.46, 38.32, 24.82, 22.78, 28.48, 24.27, 24.26, 15.52, 22.88, 15.95, 14.48, 23.83

28.72, 25.87, 22.38, 21.21, 19.90, 32.25, 24.69, 23.14, 23.81, 17.05, 24.09, 49.88

(a) Build a frequency table of the data. Does the data appear to have a normal distribution? Add relative frequency and cumulative frequency tables.

The distribution does not seem to be normal. The distribution is not symmetric There do not appear to be any outliers.

	Freq	Rel	Cum
(10,15]	1	0.0278	1
(15,20]	4	0.1111	5
(20,25]	15	0.4167	20
(25,30]	9	0.2500	29
(30,35]	2	0.0556	31
(35,40]	3	0.0833	34
(40,45]	1	0.0278	35
(45,50]	1	0.0278	36

(b) Sketch a histogram of the data. Does the histogram show a normal distribution? If not, which characteristic of normal distributions does it violate?

The histogram does not appear to be normal. It is not symmetric. It is skewed to the right.

Max Air Pollution from 2007 to 2009

2
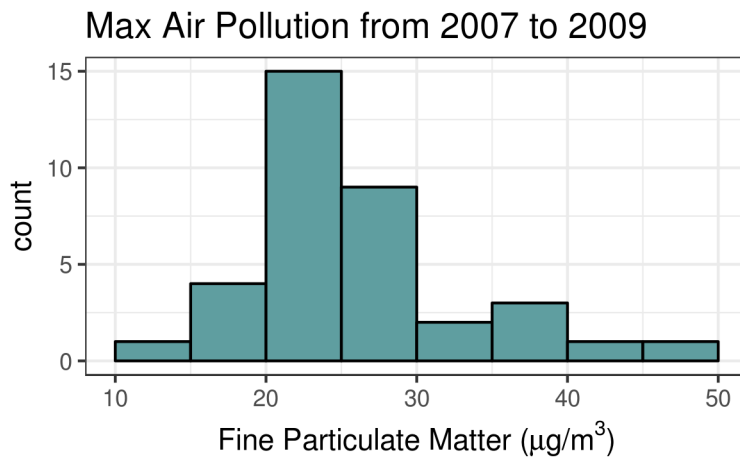
3 For each scenario below, identify some of the graphs that would be appropriate for this data. Create one appropriate graph and use it to answer the research question.

(a) The data set "faithful.csv" contains a sample of eruptions of the Old Faithful geyser in Yellowstone National Park. Length of eruption (in minutes) and waiting time until the next eruption (in minutes) are recorded. Researchers want to know if the is a relationship between eruption lengths and waiting times.

Because the question concerns the relationship between two quantitative variables, a scatterplot is most appropriate.

Old Faithful Eruption and Waiting Times
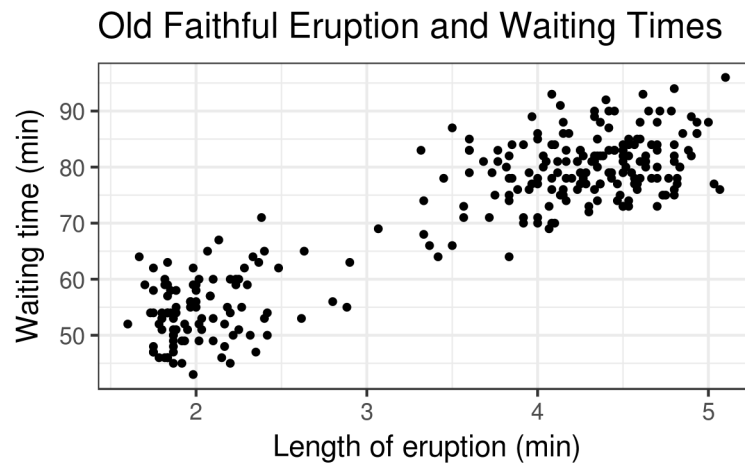
There does seem to be a relationship between length of eruption and waiting time for next eruption.

(b) The data set "hair_eye.csv" contains the hair and eye color of 592 statistics students. The math department would like to know what is the most common eye color. What is the least common eye color?

Any graph that displays counts of categorical data could be used, such as pie charts or bar graphs. However, since we want to know the most and least common eye colors, a Pareto chart would be best.

Eye Colors of Statistics Students
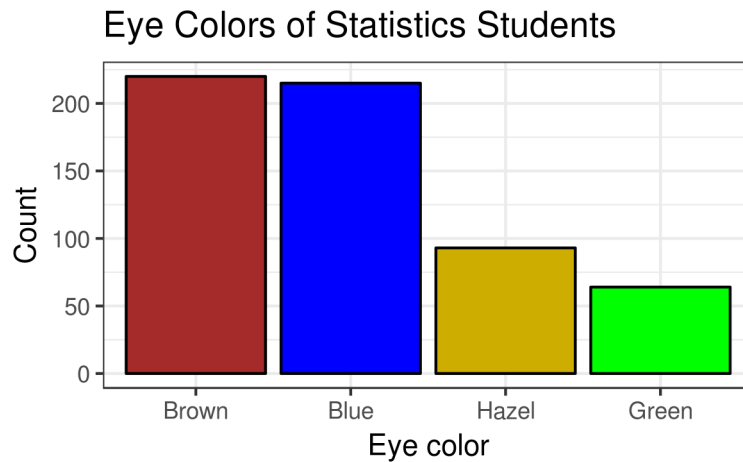
The most common eye color is brown and the least common is green.

(c) The data set "max_air_pol.csv" contains the monthly maximum fine matter particulate air pollution from 2007 to 2009. Researchers would like to know if there is a pattern in air pollution over time.

A time series is most appropriate for examining trends over time.

Time series of air pollution in MN
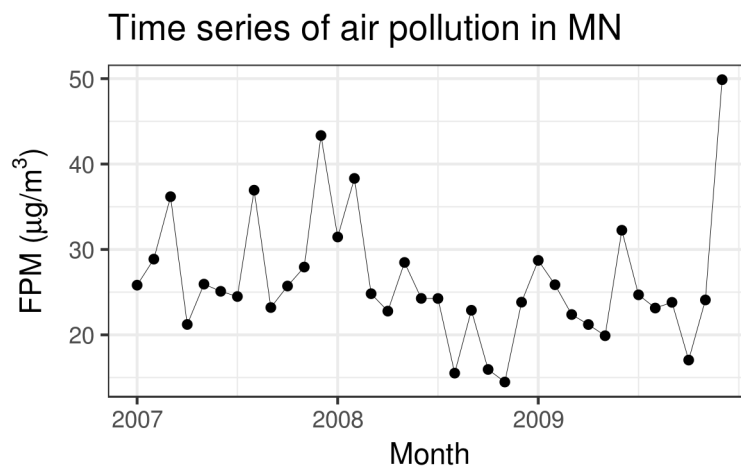
There appears to be a trend of slightly decreasing air pollution. The last value might be an outlier.

4