

Group Work - Week 4

1 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

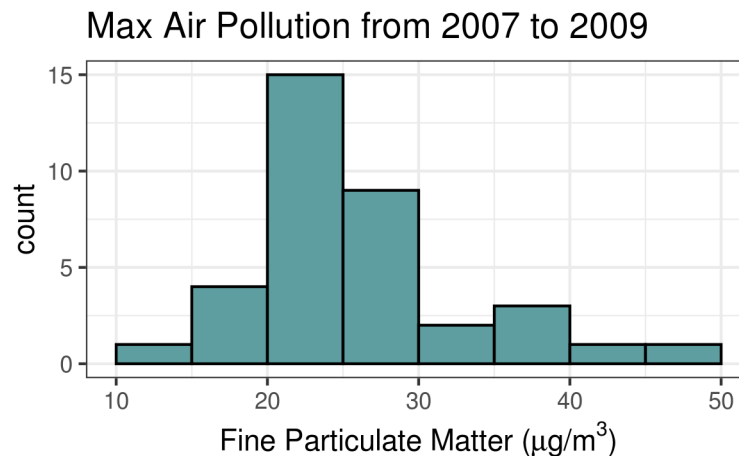
28.72, 25.87, 22.38, 21.21, 19.90, 32.25, 24.69, 23.14, 23.81, 17.05, 24.09, 49.88

- (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.**

	Frequency	Cumulative
(10,15]	1	1
(15,20]	4	5
(20,25]	15	20
(25,30]	9	29
(30,35]	2	31
(35,40]	3	34
(40,45]	1	35
(45,50]	1	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.



2 The file “max_temp_dec17.csv” on D2L contains the daily high temperatures (in F) for December.

- (a) Find the mean, median, mode and midrange of the sample. What is the most appropriate measure of center for this data?

Mean	Median	Mode	Midrange
24.65	25	22, 24, 29	24.50

Since temperatures are roughly normal and symmetric, mean is the most appropriate measure of center.

- (b) Find the range, variance and standard deviation of the sample. Be sure to include the correct units for each measure.

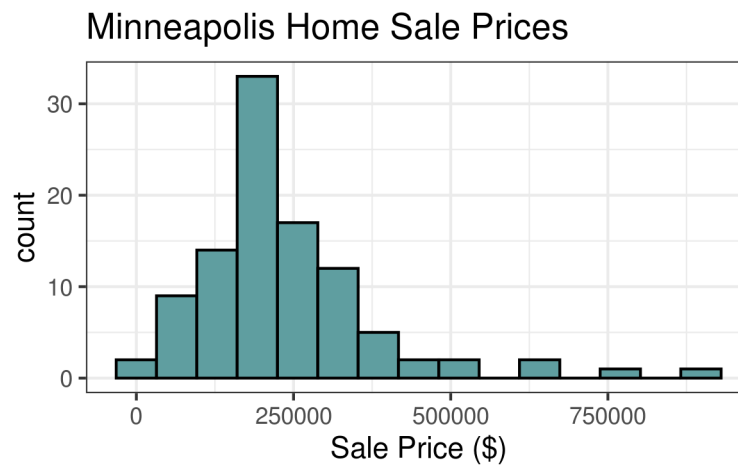
Range	Variance	SD
65 °F	252.90 (°F) ²	15.90 °F

3 The file “mpls_home_sales.csv” on D2L contains the adjusted sale prices (in dollars) of a sample of home sold in Minneapolis in 2016.

- (a) Find the mean, median, mode and midrange of the sample. What is the most appropriate measure of center for this data?

Mean	Median	Mode	Midrange
234980.59	202750.00	175000	451000.00

By looking at the histogram, we can see that the data is right skewed with a few extremely high values. Thus, median is the most appropriate measure of center.



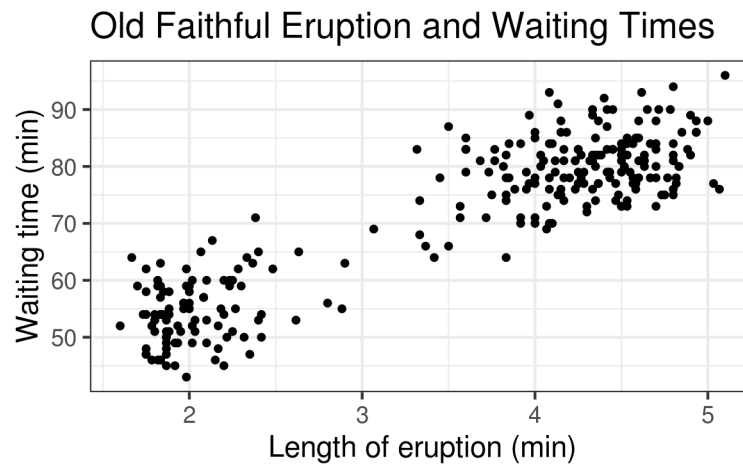
- (b) Find the range, variance and standard deviation of the sample. Be sure to include the correct units for each measure.

Range	Variance	SD
898000 \$	21434941764.24 \$ ²	146406.77 \$

4 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 there is a relationship between eruption lengths and waiting times.

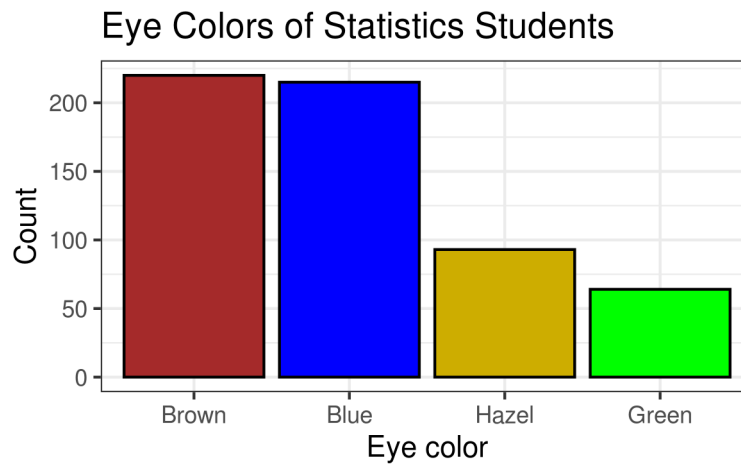
Because the question concerns the relationship between two quantitative variables, a scatterplot is most appropriate.



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.



The most common eye color is brown and the least common is green.