**Week 2 Lab: Introduction to Data**

**LATEST SUBMISSION GRADE**

1.

Question 1

Create a new data frame that includes flights headed to SFO in February, and save this data frame assfo\_feb\_flights. How many flights meet these criteria?

**1 / 1 point**



32735



68



1345



3563



2286

**Correct**

2.

Question 2

Make a histogram and calculate appropriate summary statistics for **arrival** delays of sfo\_feb\_flights. Which of the following is false?

**1 / 1 point**



The distribution has several extreme values on the right side.



The distribution is unimodal.



No flight is delayed more than 2 hours.



The distribution is right skewed.



More than 50% of flights arrive on time or earlier than scheduled.

**Correct**

3.

Question 3

Calculate the median and interquartile range for arr\_delays of flights in the sfo\_feb\_flights data frame, grouped by carrier. Which carrier has the highest IQR of arrival delays?

**1 / 1 point**



Delta and United Airlines



JetBlue Airways



Frontier Airlines



Virgin America



American Airlines

**Correct**

4.

Question 4

Considering the data from all the NYC airports, which month has the highest average departure delay?

**1 / 1 point**



January



March



December



October



July

**Correct**

5.

Question 5

Which month has the highest median departure delay from an NYC airport?

**1 / 1 point**



March



July



December



January



October

**Correct**

6.

Question 6

Is the mean or the median a more reliable measure for deciding which month(s) to avoid flying if you really dislike delayed flights, and why?

**1 / 1 point**



Mean would be more reliable as the distribution of delays is symmetric.



Median would be more reliable as the distribution of delays is skewed.



Mean would be more reliable as it gives us the true average.



Median would be more reliable as the distribution of delays is symmetric.



Both give us useful information.

**Correct**

7.

Question 7

If you were selecting an airport simply based on on time departure percentage, which NYC airport would you choose to fly out of?

**1 / 1 point**



EWR



LGA



JFK

**Correct**

8.

Question 8

Mutate the data frame so that it includes a new variable that contains the average speed, avg\_speed traveled by the plane for each journey (in mph). What is the tail number of the plane with the fastest avg\_speed? **Hint:** Average speed can be calculated as distance divided by number of hours of travel, and note that air\_time is given in minutes. If you just want to show the avg\_speed and tailnum and none of the other variables, use the select function at the end of your pipe to select just these two variables with select(avg\_speed, tailnum). You can google this tail number to find out more about the aircraft.

**1 / 1 point**



N947UW



N666DN



N779JB



N755US



N959UW

**Correct**

9.

Question 9

Make a scatterplot of avg\_speed vs. distance. Which of the following is true about the relationship between average speed and distance.

**1 / 1 point**



There is an overall positive association between distance and average speed.



The distribution of distances are uniform over 0 to 5000 miles.



The relationship is linear.



There are no outliers.



As distance increases the average speed of flights decreases.

**Correct**

10.

Question 10

Suppose you define a flight to be “on time” if it gets to the destination on time or earlier than expected, regardless of any departure delays. Mutate the data frame to create a new variable called arr\_type with levels "on time" and "delayed" based on this definition. Also mutate to create a new variable called dep\_type with levels "on time" and "delayed" depending on the flight was delayed for fewer than 5 minutes or 5 minutes or more, respectively. In other words, if arr\_delay is 0 minutes or fewer, arr\_type is "on time". If dep\_delay is less than 5 minutes, dep\_type is "on time". Then, determine the on time arrival percentage based on whether the flight departed on time or not. What fraction of flights that were "delayed" departing arrive "on time"? (Enter the answer in decimal point, like 0.xx)

**1/1**

18.33