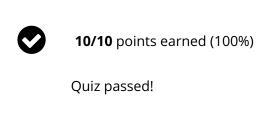
Week 2 Lab: Introduction to Data



Back to Week 2



1/1 points

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?

32735

68

Correct Response

1345

3563

2286



1/1

points

2.

delays of sfo_feb_flights. Which of the following is false?			
0	No flight is delayed more than 2 hours.		
Correct Response			
0	The distribution has several extreme values on the right side.		
0	The distribution is right skewed.		
0	The distribution is unimodal.		
0	More than 50% of flights arrive on time or earlier than scheduled.		
sfo_fek	1/1 points ate the median and interquartile range for arr_delays of flights in the o_flights data frame, grouped by carrier. Which carrier has the highest arrival delays?		
0	JetBlue Airways		
0	Frontier Airlines		
0	American Airlines		
0	Virgin America		
0	Delta and United Airlines		
Correct Response			



1/1 points

4.

Which month has the highest average departure delay from an NYC airport? July **Correct Response** January March October December 1/1 points 5. Which month has the highest median departure delay from an NYC airport? October January July December **Correct Response** March 1/1 points 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?

	symmetric.
0	Median would be more reliable as the distribution of delays is symmetric.
0	Median would be more reliable as the distribution of delays is skewed.
Cor	rect Response
0	Mean would be more reliable as it gives us the true average.
0	Both give us useful information.
	1 / 1 points were selecting an airport simply based on on time departure
-	
If you percer	points were selecting an airport simply based on on time departure ntage, which NYC airport would you choose to fly out of?
If you percer	points were selecting an airport simply based on on time departure ntage, which NYC airport would you choose to fly out of? LGA

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.

about	the aircraft.		
0	N779JB		
0	N959UW		
0	N755US		
0	N666DN		
Corr	ect Response		
0	N947UW		
~	1/1 points		
	a scatterplot of avg_speed vs. distance. Which of the following is true the relationship between average speed and distance.		
0	The relationship is linear.		
0	As distance increases the average speed of flights decreases.		
0	The distribution of distances are uniform over 0 to 5000 miles.		
0	There is an overall positive association between distance and average speed.		
Correct Response			

There are no outliers.

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. Then, determine the on time arrival percentage based on whether the flight departed on time or not. What proportion of flights that were "delayed" departing arrive "on time"? (answer should be in the form 0.## where ## is between 2 and 7 decimal places, inclusive)

0.1833639

Correct Response

