Homework 2

By signing my name, I agree to abide by the Stern Code of Conduct _____

Question 1

This question follows **Section 3.3** in the ModernDive book linked on the course page: http://moderndive.com/3-viz.html

- Load the libraries `dplyr, ggplot2, nycflights13`.
- Run the command `?flights` and read the help documentation to learn about what variables are included in this data frame.
- Run the command `head(flights)` and look at the first few rows of data.
- Use the `table()` function to find counts of the `carrier` variable.

a. How many flights are there from Hawaiian Airlines? Which airline had the most number of flights? Write your answers here.

- Following the example in the book, use `filter()` to create a data frame with all Hawaiian Airlines flights (*instead of Alaskan Airlines*). Use this data frame to complete the parts below.
- Create a histogram of the `air_time` variable using `ggplot()` and `geom_histogram()`.
- (Remember you can use `+ theme_minimal()` to make the plot use less ink)
- Use the `ggsave()` command to save your plot as a pdf file named "air_time_hist.pdf"

b. Upload your plot on Classes under the File Exchange tab.

c. If you wanted to describe a histogram using summary statistics, which summaries would you use, and why?

d. What are the mean and standard deviation of the `air_time` variable for these Hawaiian Airline flights?

e. What percent of these lie within one, two, and three standard deviation of the mean? Include 4 digits of accuracy for each, e.g. 12.34%

f. Save all of the `R` code you used to do this problem in a file called "hw1.R" and upload it on Classes.

Question 2

a. Consider a game involving tossing a coin three times, and use set notation (e.g. { item1, item2, ... }) to answer the following: What is the sample space S? What is the event E_1 = "the first toss is H"? What is the intersection of E_1 and E_2 = "the second toss is H"?

b. A 6-sided die is rolled 10 times. What is the probability it comes up six every time? What is the probability of no sixes? What is the probability every roll is less or equal to 5?

c. Every time I roll a 20-sided die there is a 1/20 chance it lands on 20. So if I roll it 10 times, the probability that at least one of these will be a 20 is 10/20 = 1/2. True or false? Explain your answer.

d. What is the equation for the conditional probability of E_2 given E_1? Suppose $P(E_1) > P(E_2)$ and $P(E_1 \cap E_2) > 0$. Which is larger, $P(E_2|E_1)$ or $P(E_1|E_2)$, or is there not enough information to tell? Explain, and also draw a Venn diagram as an example.

e. The mortality rates from a hypothetical drug trial are in the table below. Is survival independent from treatment/placebo, adherence/non-adherence, both of these, or neither? Explain.

	Treatment	Placebo
Adherers	15%	15%
Non-adherers	25%	25%
Total	20%	20%