Project 1: Getting Familiar with SAS

Name: Kameron Kyger

Part 1: Descriptive Statistics

Questions that you must answer are highlighted in yellow. For this exercise, we will be using a built-in dataset in SAS to answer the following questions about cars.

First, create a temporary dataset named “two” using the following code:

```
data two;
set sashelp.cars;
run;
```

Using PROC MEANS. Find the mean and standard deviations of the following variables (Please paste your SAS Output table below):

- Invoice
- Horsepower
- Weight

The SAS System

The MEANS Procedure

Variable	Label	Mean	Std Dev
Invoice		30014.70	17642.12
Horsepower		215.8855140	71.8360316
Weight	Weight (LBS)	3577.95	758.9832146

There are different locations of origin and drive train types included in this dataset. Let’s find out how many and what they are using PROC FREQ. Run the following code:

```
proc freq data=two;
table origin drivetrain / missing;
run;
```

Paste your output table.

The FREQ Procedure

Origin	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Asia	158	36.92	158	36.92
Europe	123	28.74	281	65.65
USA	147	34.35	428	100.00

DriveTrain	Frequency	Percent	Cumulative Frequency	Cumulative Percent
All	92	21.50	92	21.50
Front	226	52.80	318	74.30
Rear	110	25.70	428	100.00

How many locations of origin are included in this dataset? Drive trains? 3 Origins and 3 Drive Trains

What are the locations? Asia, Europe, and USA

Now that we know that there are multiple drive trains – use the information from Practice and Participation 1 to find the mean weight and standard deviation based on drive train.

Please complete the following table. Add or delete rows as necessary.

DriveTrain	Mean	Standard Deviation
All	4171.37	886.7474549
Front	3308.64	651.1054267
Rear	3634.96	523.6399584

Suppose we are interested in examining horsepower in terms of < 300 and ≥ 300. Let's create a new variable to dichotomize horsepower. Run the following code:

```
data two;
set two;
if .< horsepower < 300 then HP300 = 0;
else if horsepower > 55 then HP300 = 1;
run;
```

To see what this code accomplished, please use this code (

Obs	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length	HP300
1	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6	265	17	23	4451	106	189	0
2	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4	200	24	31	2778	101	172	0
3	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4	200	22	29	3230	105	183	0
4	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6	270	20	28	3575	108	186	0
5	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6	225	18	24	3880	115	197	0
6	Acura	3.5 RL w/Navigation 4dr	Sedan	Asia	Front	$46,100	$41,100	3.5	6	225	18	24	3893	115	197	0
7	Acura	NSX coupe 2dr manual S	Sports	Asia	Rear	$89,765	$79,978	3.2	6	290	17	24	3153	100	174	0
8	Audi	A4 1.8T 4dr	Sedan	Europe	Front	$25,940	$23,508	1.8	4	170	22	31	3252	104	179	0
9	Audi	A41.8T convertible 2dr	Sedan	Europe	Front	$35,940	$32,506	1.8	4	170	23	30	3638	105	180	0
10	Audi	A4 3.0 4dr	Sedan	Europe	Front	$31,840	$28,846	3.0	6	220	20	28	3462	104	179	0

):

```
proc print data=two (obs=10);
run;
```

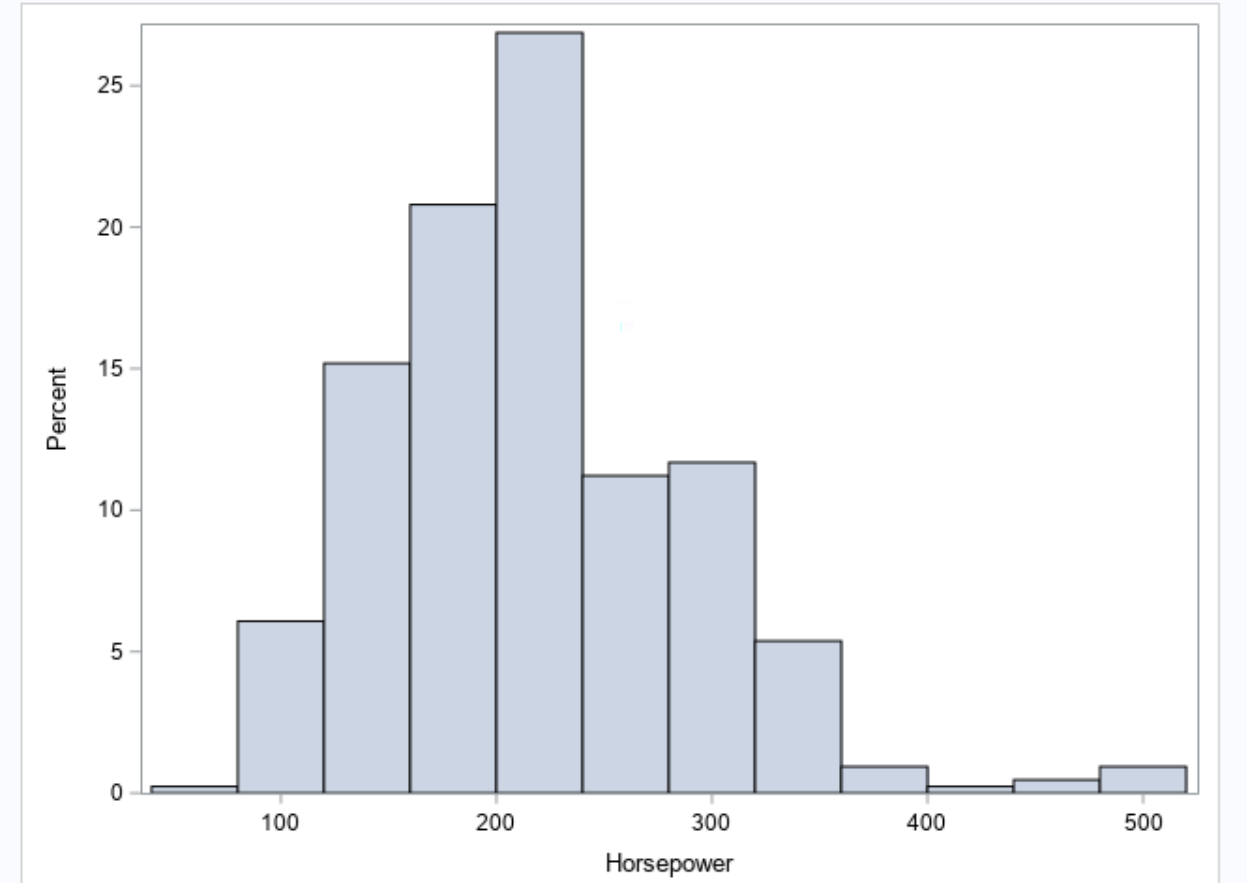
Investigate the number of each type above and below the cut off of 300HP. Run the following code:

```
proc freq data=two;
table type*HP300 / missing;
run;
```

Determine the percentage of all cars above and below 300HP. (Above: 85.75% Below: 14.25%)

Use your code from Practice and Participation 1 to create a histogram of Horsepower.

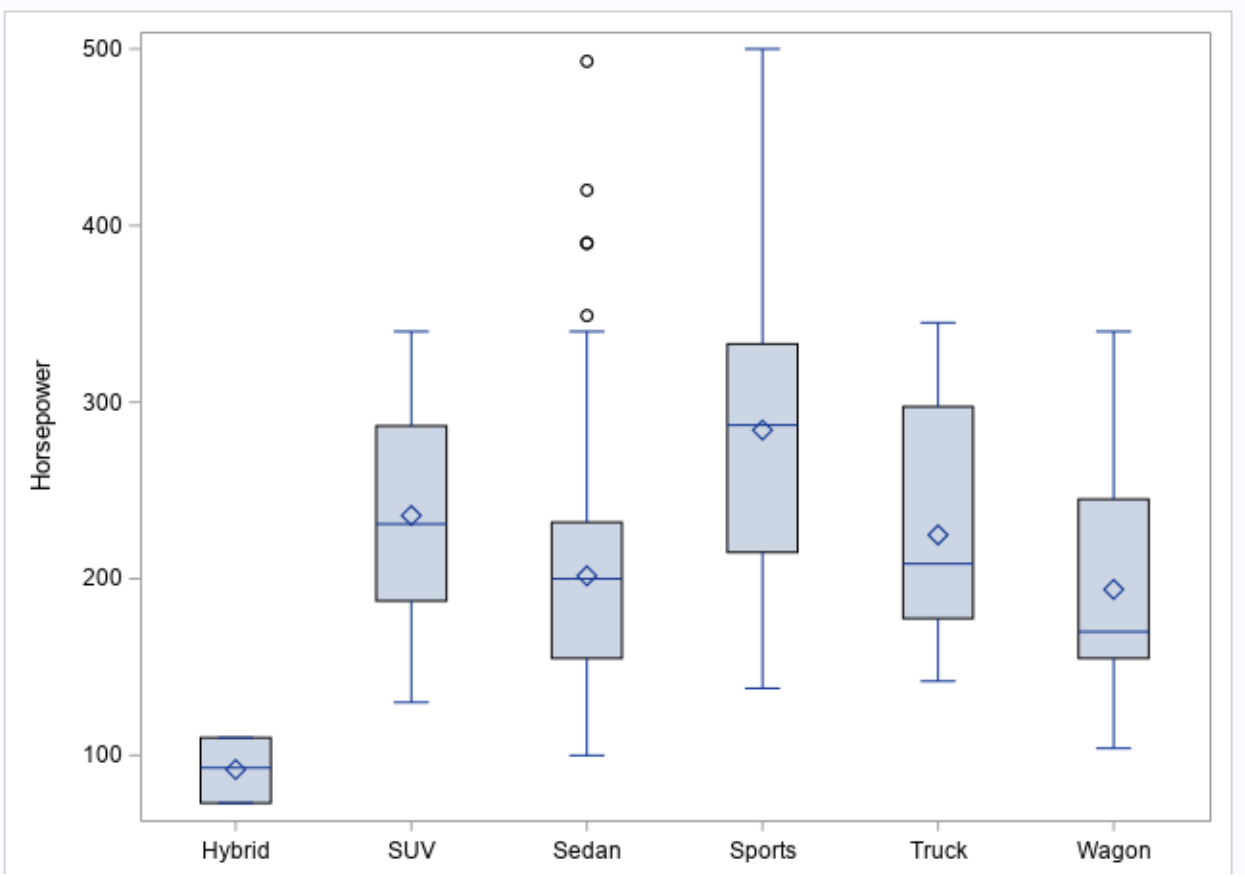
Take a screenshot of the histogram (Windows: use the Snipping Tool; Mac: press command + shift + 4) and include it below:



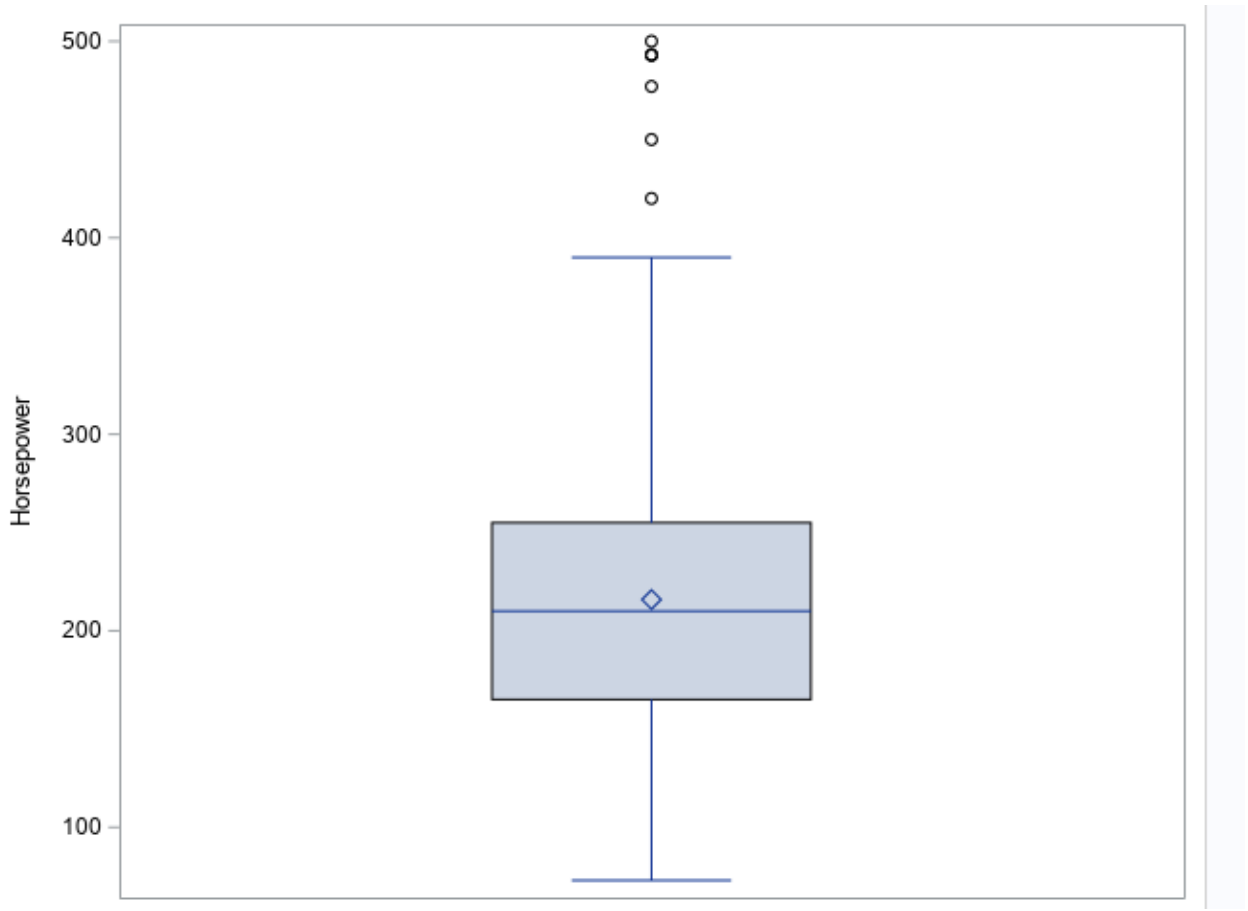
Create a box plot of Horsepower, split by type of vehicle using the following code:

```
proc sgplot data=two;
vbox horsepower / category=type;
run;
```

Take a screenshot of the box plot and include it below:



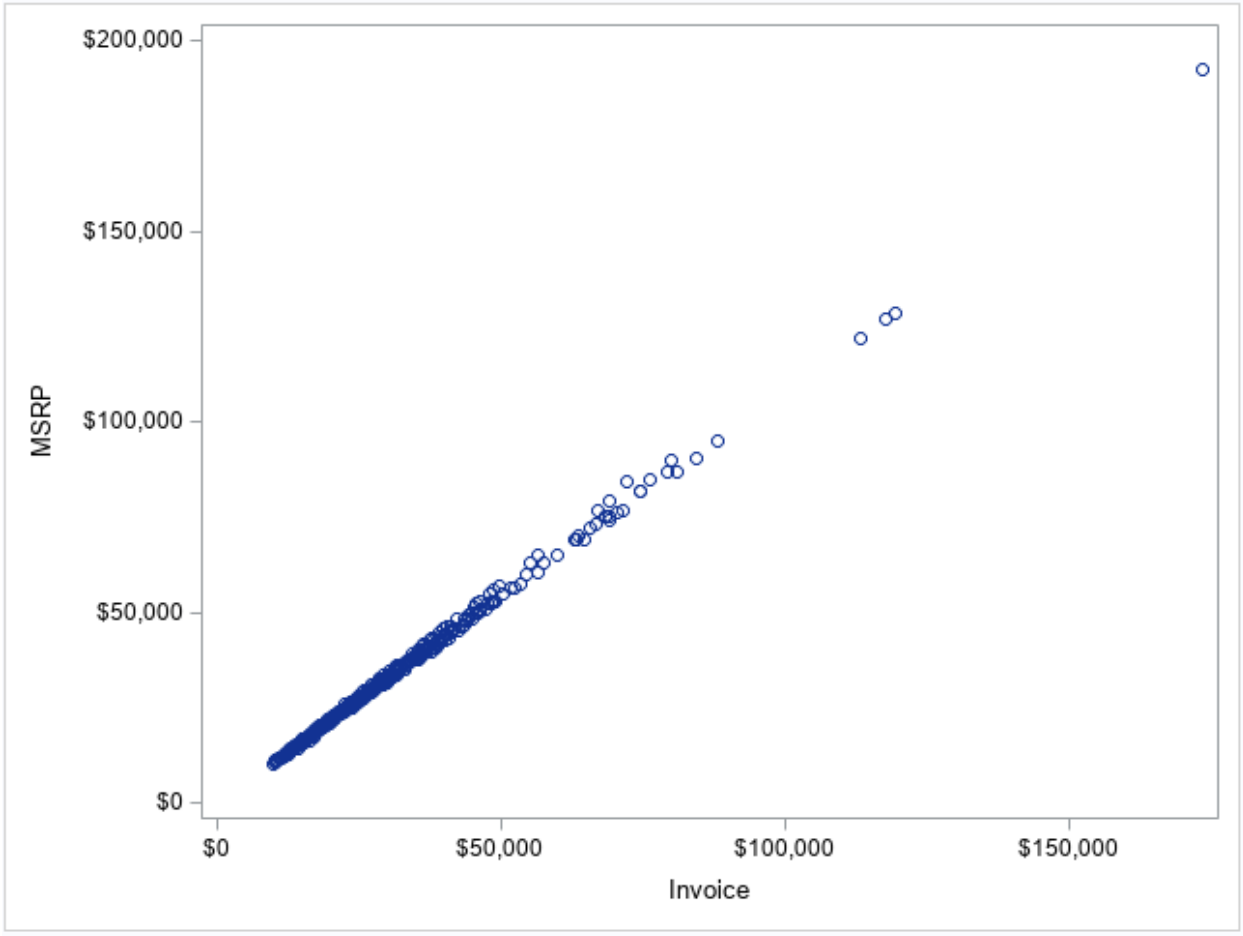
Create a single boxplot of horsepower. Include it below.



Finally, let's create a scatterplot of invoice and msrp using the following code:

```
proc sgplot data=two;
scatter x = invoice y = msrp;
run;
```

Take a screenshot of the scatter plot and include it below:



Does it appear that invoice and msrp are related? If we were to find the correlation, do you expect it to be positive or negative?

It does appear that the two are related. I expect it to be a positive correlation

Part 2: Probability

Calculate the following probabilities using the SAS codes as appropriate:

```
data exbinom;
binom1 = pdf('BINOM',x,p,n); run;
proc print data=exbinom; run;

data exbinom;
binom1 = cdf('BINOM',x,p,n); run;
proc print data=exbinom; run;

data exnorm;
norm1 = probnorm(Z); run;
proc print data=exnorm; run;

data exnorm;
```

```
norm1 = probnorm((x-mu)/sigma); run;
proc print data=exnorm; run;
```

1. According to a report of national health statistics, the weight of male babies in the US is normally distributed with a mean of 11.5lb and a standard deviation of 2.7lb. We select a random male baby under 2 months.
    a. Is this a binomial or normal probability experiment? Normal
    b. What is the probability that the baby weighs less than 8lb? 9.7437%
    c. What is the probability that the baby weighs more than 6lb? 97.918%
    d. What is the probability that the baby weighs between 11.2lb and 14lb? 36.699%
    e. What is the probability that the baby weighs at most 12lb? 57.346%

2. Of all the registered cars in Colorado, 89% pass the annual emissions inspection. We select 32 random cars being inspected for emissions.
    a. Is this a binomial or normal probability experiment? Binomial
    b. What is the probability that exactly 25 cars will pass the inspection? 3.5612%
    c. What is the probability that at most 25 cars will pass the inspection? 5.5592%
    d. What is the probability that more than 20 cars will pass the inspection? 99.998%
    e. What is the probability that between 14 and 21 cars will pass the inspection? 0.0403127%