

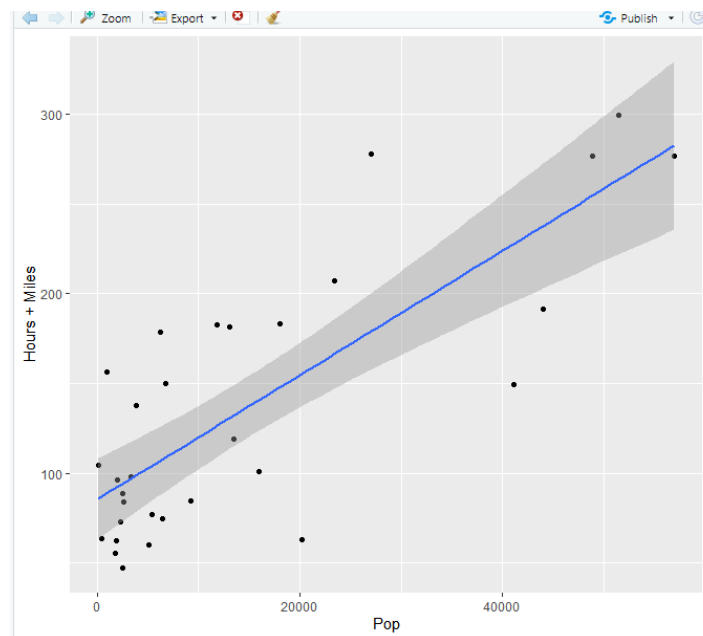
Name(s): Caleb Lich, Jacob Blovins, Nathan Hynek, Michael Steets

8.2 (Linear Regression)

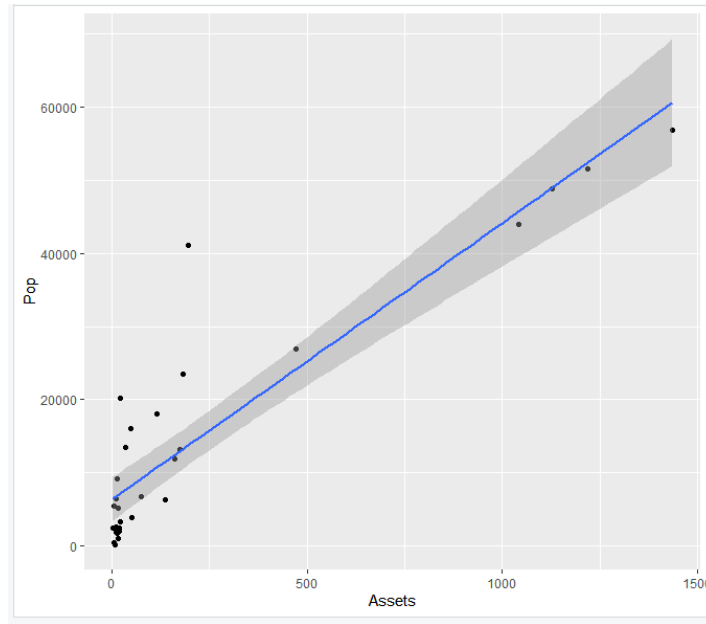
Questions relating to Airline Costs dataset

Use a linear regression model to predict the number of customers each airline serves from its length of flight and daily flight time per plane. Next, build another regression model to predict the total assets of an airline from the customers served by the airline. Do you have any insight about the data from the last two regression models?

- a. `ggplot(planes, aes(X = Pop, Y = Hours + Miles)) + geom_point() + stat_smooth(method = "lm")`



- b. `ggplot(planes, aes(x = Assets, y = Pop))+ geom_point() + stat_smooth(method = "lm")`



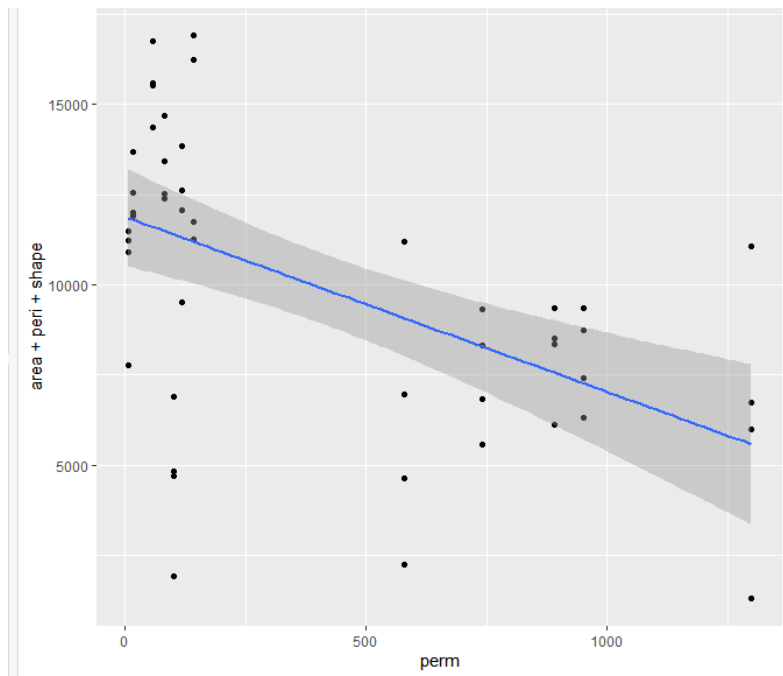
INSIGHT: For the first graph, we'd assume that people would like to take shorter flights than longer ones. For the second graph, we'd assume that people try to take advantage of cheaper flights which is why the most assets are coming from the lower area. Finally, we assumed that the cheaper flights and the shorter flights are connected in that most people choose shorter flights because they are cheaper.

8.3 (Gradient descent)

Questions relating to BP Research dataset

First, create a linear model and check if the perm has linear relationship with the remaining three attributes. Next, use the gradient descent algorithm to find the optimal intercept and gradient for the dataset.

- a. `ggplot(RockData, aes(x = perm, y = area))+ geom_point() + stat_smooth(method = "lm")`



```
b. > set.seed(1)
> df=rapply(RockData, scale, c("numeric", "integer"), how = "replace")
> df
```

	...1[,1]	area[,1]	peri[,1]
	<dbl>	<dbl>	<dbl>
1	-1.68	-0.819	0.0766
2	-1.61	-0.0692	0.845
3	-1.54	0.138	0.872
4	-1.46	0.0612	0.829
5	-1.39	0.281	0.885
6	-1.32	0.295	0.928
7	-1.25	0.799	1.16
8	-1.18	0.381	1.16
9	-1.11	0.449	0.698
10	-1.04	-0.284	0.291

INSIGHT: The graph doesn't track with the linear model so gradient track is needed. We predict that the smaller the rock, the bigger the permeability.