

Problem 1: James 2.4, Exercise 8

Part a

```
> college <- read.csv(file="datasets/College.csv")
> head(college)
```

		X	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books
1	Abilene Christian University	Yes	1660	1232	721	23	52	2885	537	7440	3300	450	
2	Adelphi University	Yes	2186	1924	512	16	29	2683	1227	12280	6450	750	
3	Adrian College	Yes	1428	1097	336	22	50	1036	99	11250	3750	400	
4	Agnes Scott College	Yes	417	349	137	60	89	510	63	12960	5450	450	
5	Alaska Pacific University	Yes	193	146	55	16	44	249	869	7560	4120	800	
6	Albertson College	Yes	587	479	158	38	62	678	41	13500	3335	500	

	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
1	2200	70	78	18.1	12	7041	60
2	1500	29	30	12.2	16	10527	56
3	1165	53	66	12.9	30	8735	54
4	875	92	97	7.7	37	19016	59
5	1500	76	72	11.9	2	10922	15
6	675	67	73	9.4	11	9727	55

Part b

	row.names	Private	Apps	Accept	Enroll	Top10perc	Top25perc
1	Abilene Christian University	Yes	1660	1232	721	23	52
2	Adelphi University	Yes	2186	1924	512	16	29
3	Adrian College	Yes	1428	1097	336	22	50
4	Agnes Scott College	Yes	417	349	137	60	89
5	Alaska Pacific University	Yes	193	146	55	16	44
6	Albertson College	Yes	587	479	158	38	62
7	Albertus Magnus College	Yes	353	340	103	17	45
8	Albion College	Yes	1899	1720	489	37	68
9	Albright College	Yes	1038	839	227	30	63
10	Alderson-Broaddus College	Yes	582	498	172	21	44

Part c

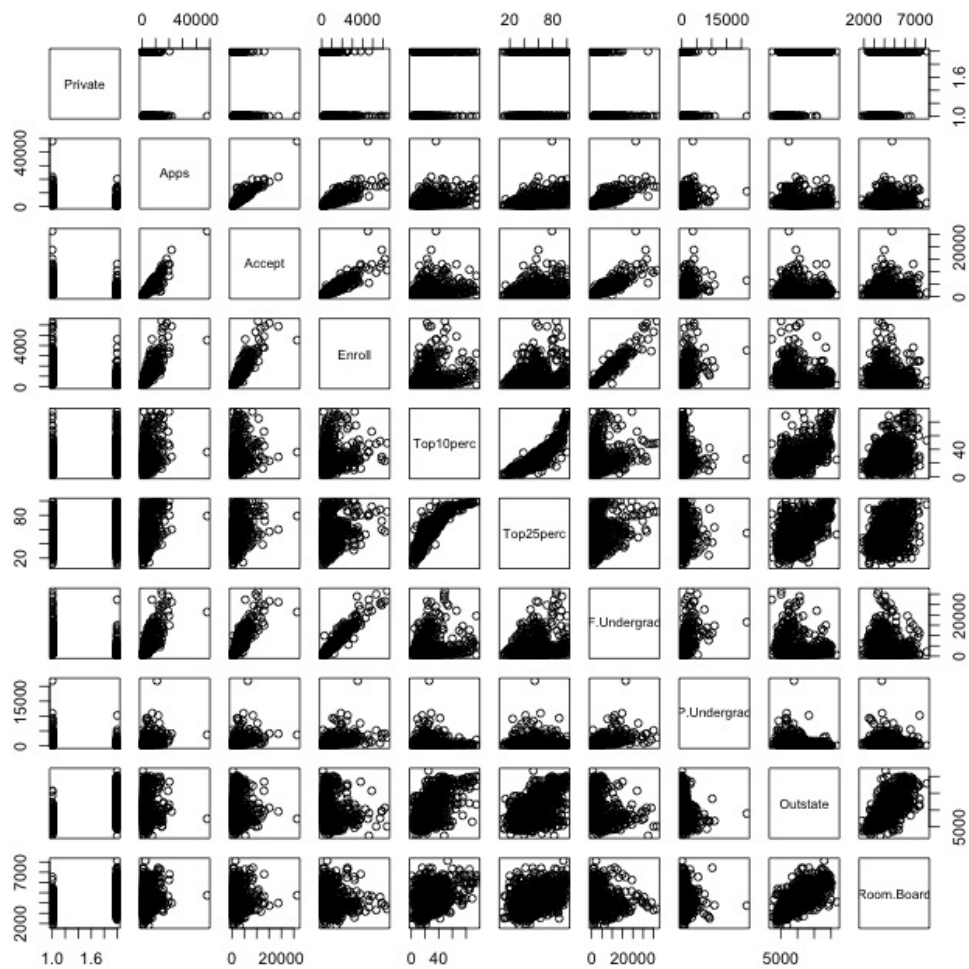
Part i

```
> summary(college)
```

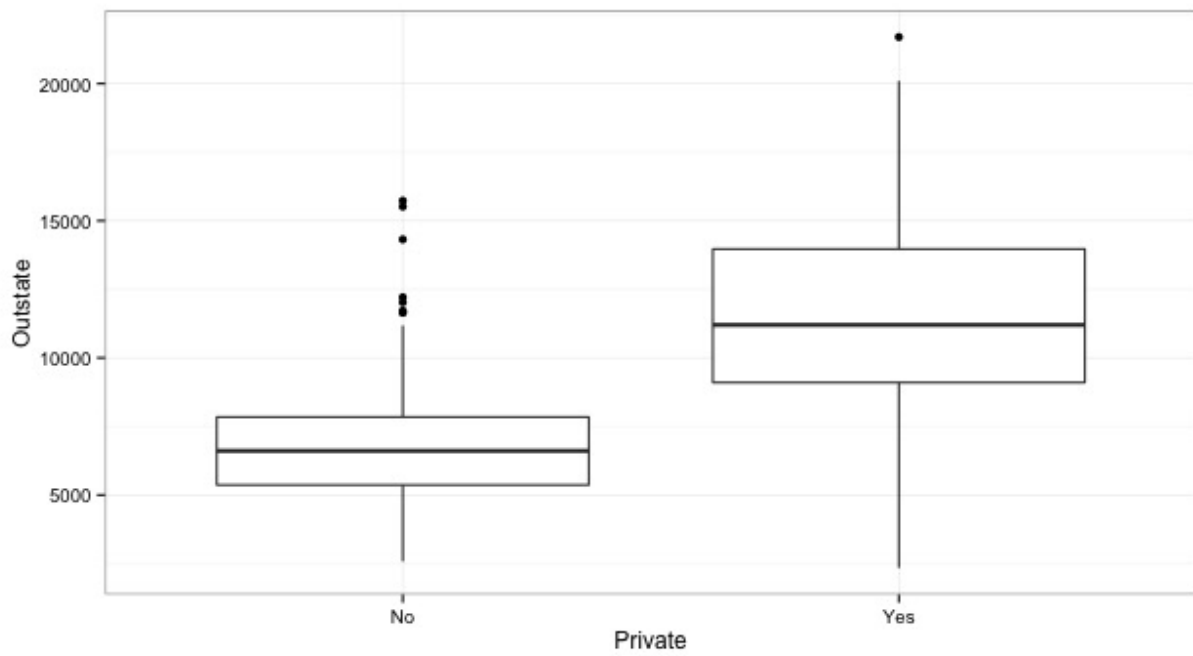
Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate
No :212	Min. : 81	Min. : 72	Min. : 35	Min. : 1.00	Min. : 9.0	Min. : 139	Min. : 1.0	Min. : 2340
Yes:565	1st Qu.: 776	1st Qu.: 684	1st Qu.: 242	1st Qu.:15.00	1st Qu.: 41.0	1st Qu.: 992	1st Qu.: 95.0	1st Qu.: 7320
	Median : 1558	Median : 1110	Median : 434	Median :23.00	Median : 54.0	Median : 1707	Median : 353.0	Median : 9990
	Mean : 3002	Mean : 2019	Mean : 780	Mean :27.56	Mean : 55.8	Mean : 3700	Mean : 855.3	Mean :10441
	3rd Qu.: 3624	3rd Qu.: 2424	3rd Qu.: 902	3rd Qu.:35.00	3rd Qu.: 69.0	3rd Qu.: 4005	3rd Qu.: 967.0	3rd Qu.:12925
	Max. :48094	Max. :26330	Max. :6392	Max. :96.00	Max. :100.0	Max. :31643	Max. :21836.0	Max. :21700

Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
Min. :1780	Min. : 96.0	Min. : 250	Min. : 8.00	Min. : 24.0	Min. : 2.50	Min. : 0.00	Min. : 3186	Min. : 10.00
1st Qu.:3597	1st Qu.: 470.0	1st Qu.: 850	1st Qu.: 62.00	1st Qu.: 71.0	1st Qu.:11.50	1st Qu.:13.00	1st Qu.: 6751	1st Qu.: 53.00
Median :4200	Median : 500.0	Median :1200	Median : 75.00	Median : 82.0	Median :13.60	Median :21.00	Median : 8377	Median : 65.00
Mean :4358	Mean : 549.4	Mean :1341	Mean : 72.66	Mean : 79.7	Mean :14.09	Mean :22.74	Mean : 9660	Mean : 65.46
3rd Qu.:5050	3rd Qu.: 600.0	3rd Qu.:1700	3rd Qu.: 85.00	3rd Qu.: 92.0	3rd Qu.:16.50	3rd Qu.:31.00	3rd Qu.:10830	3rd Qu.: 78.00
Max. :8124	Max. :2340.0	Max. :6800	Max. :103.00	Max. :100.0	Max. :39.80	Max. :64.00	Max. :56233	Max. :118.00

Part ii

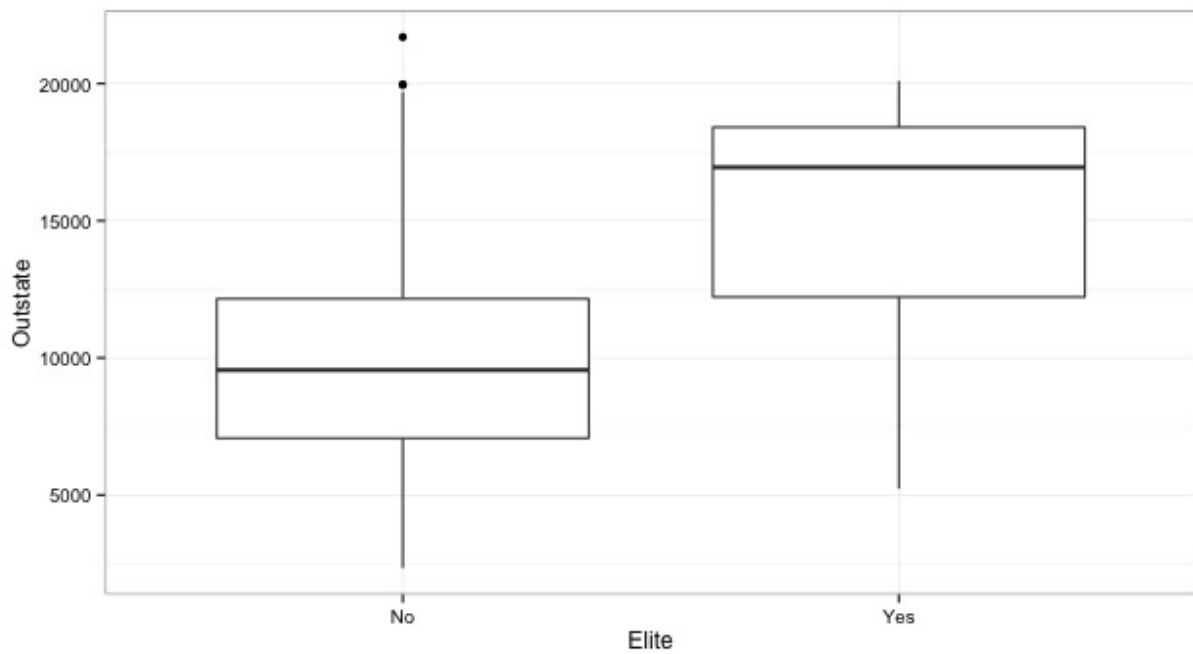


Part iii

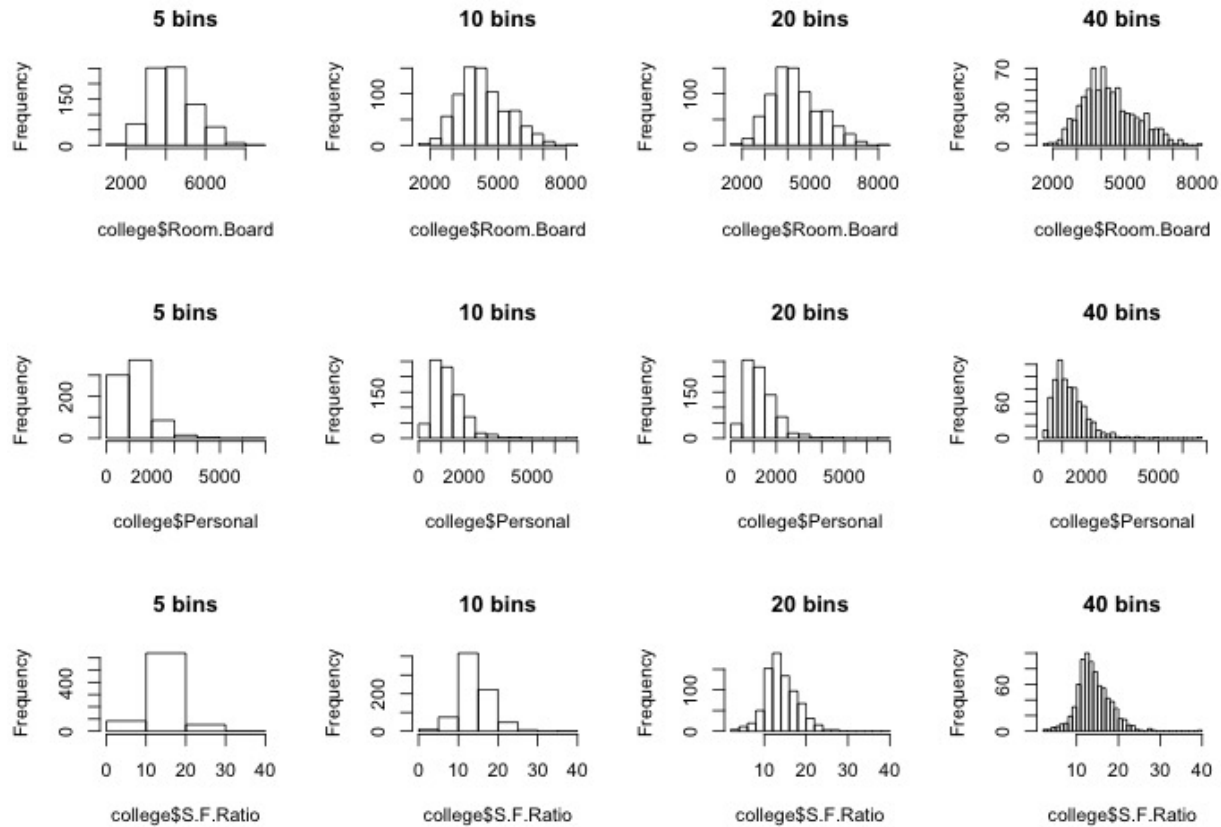


Part iv

There are 78 colleges categorized as "Elite".

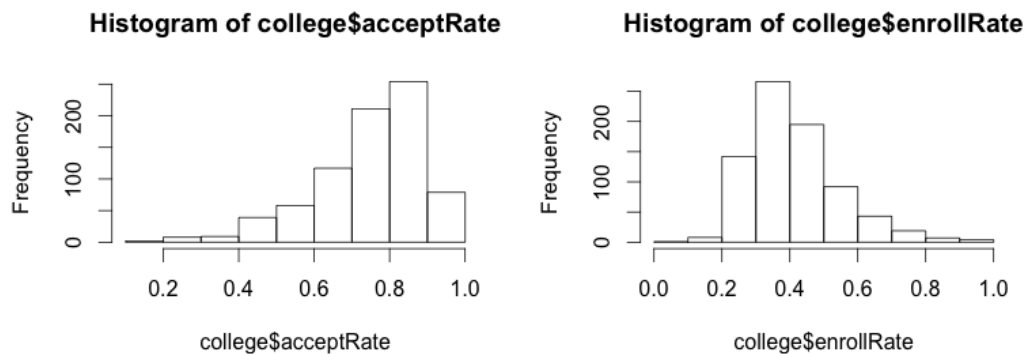


Part v



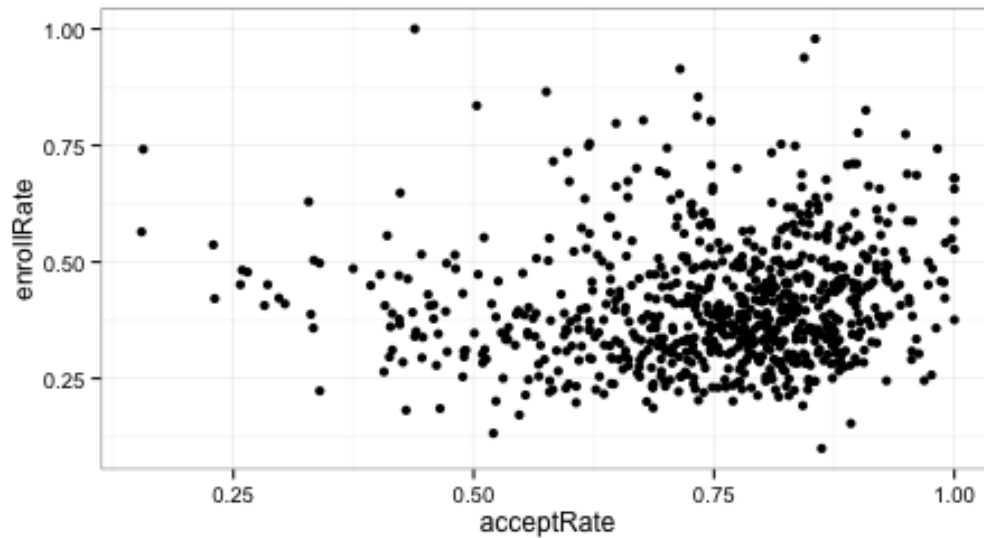
Part vi

I chose to explore acceptance and enrollment rates, with acceptance rate defined as the number of students accepted per application received (`college$Accept/college$Apps`), and enrollment rate defined as the number of students enrolled per applicant accepted (`college$Enroll/college$Accept`).



Somewhat surprisingly, there's little correlation between acceptance and enrollment rate.

```
> cor(college$acceptRate, college$enrollRate)
[1] 0.0824304
```

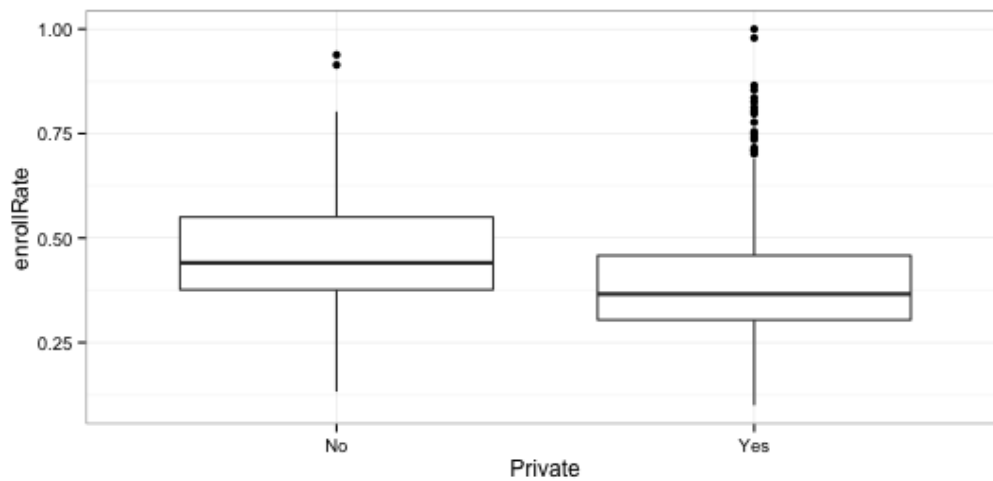
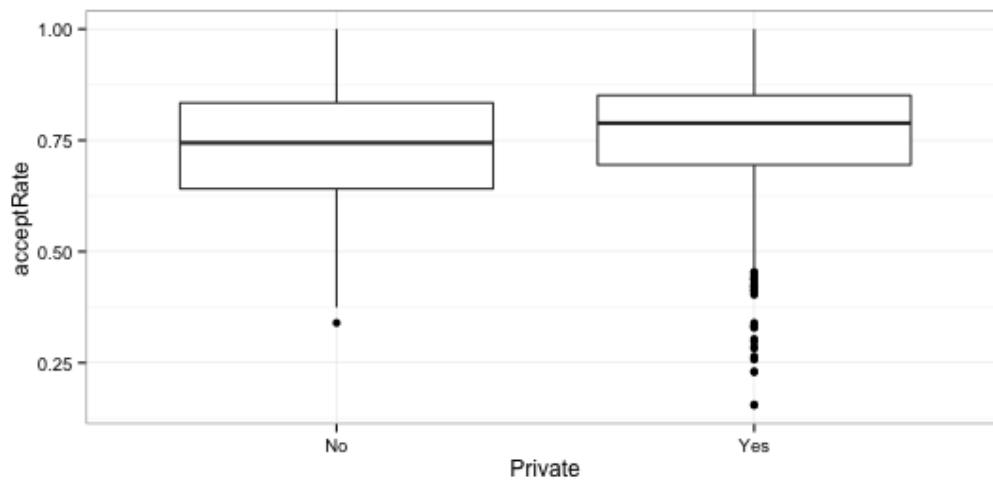


I also looked at summary statistics by public vs private schools.

```
> college %>%
+   group_by(Private) %>%
+   summarize(min.acceptRate=min(acceptRate),
+             median.acceptRate=median(acceptRate),
+             mean.acceptRate=mean(acceptRate),
+             max.acceptRate=max(acceptRate),
+             min.enrollRate=min(enrollRate),
+             median.enrollRate=median(enrollRate),
+             mean.enrollRate=mean(enrollRate),
+             max.enrollRate=max(enrollRate)) %>%
+   as.data.frame()
```

	Private	min.acceptRate	median.acceptRate	mean.acceptRate	max.acceptRate	min.enrollRate
1	No	0.3397060	0.7443387	0.7265305	1	0.13242009
2	Yes	0.1544863	0.7885653	0.7545812	1	0.09975397

	median.enrollRate	mean.enrollRate	max.enrollRate
1	0.4405908	0.4620216	0.9382716
2	0.3660934	0.3932510	1.0000000



Problem 2: [James](#) 2.4, Exercise 9

Part a

Quantitative predictors: mpg, cylinders, displacement, horsepower, weight, acceleration, year

Qualitative predictors: origin, name

Part b

	statistic	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin	name
1	min	9.0	3	68	46	1613	8.0	70	NA	NA
2	max	46.6	8	455	230	5140	24.8	82	NA	NA

Part c

	statistic	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin
1	mean	23.445918	5.471939	194.412	104.46939	2977.5842	15.541327	75.979592	NA
2	sd	7.805007	1.705783	104.644	38.49116	849.4026	2.758864	3.683737	NA

Part d

	statistic	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin
1	min	11.000000	3.000000	68.00000	46.00000	1649.0000	8.500000	70.000000	NA
2	max	46.600000	8.000000	455.00000	230.00000	4997.0000	24.800000	82.000000	NA
3	mean	24.404430	5.373418	187.24051	100.72152	2935.9715	15.726899	77.145570	NA
4	sd	7.867283	1.654179	99.67837	35.70885	811.3002	2.693721	3.106217	NA

Part e

MPG generally increased between 1970 and 1982.

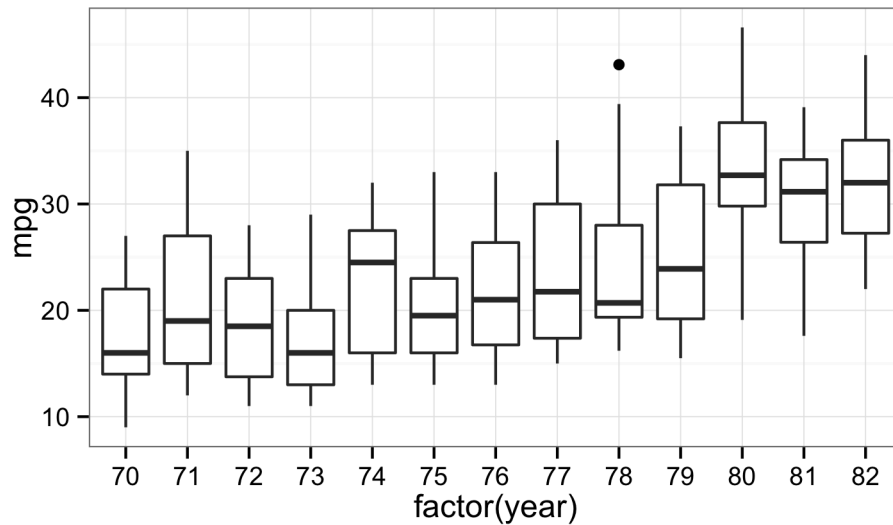


Figure 1: MPG vs year

As the number of cylinders increases, MPG generally decreases.

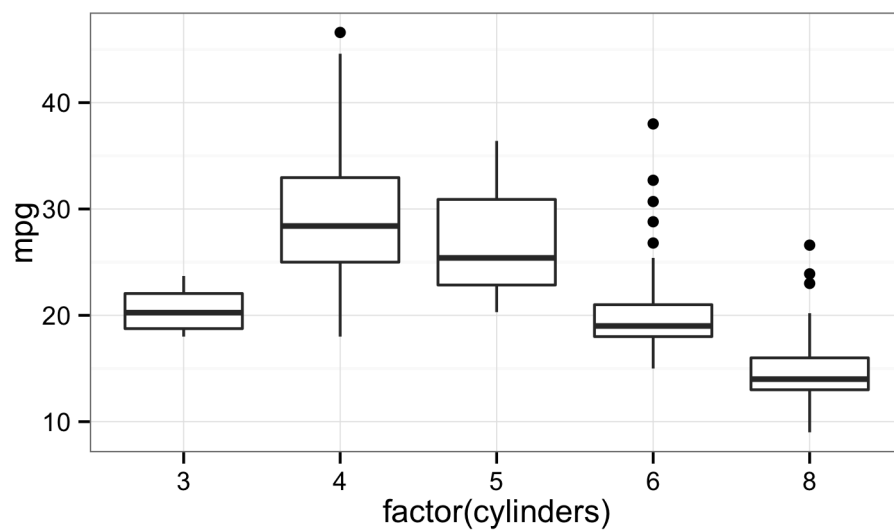
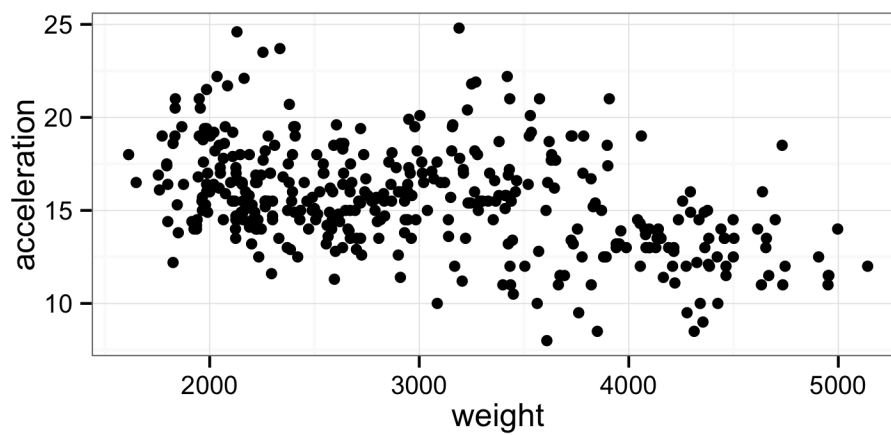
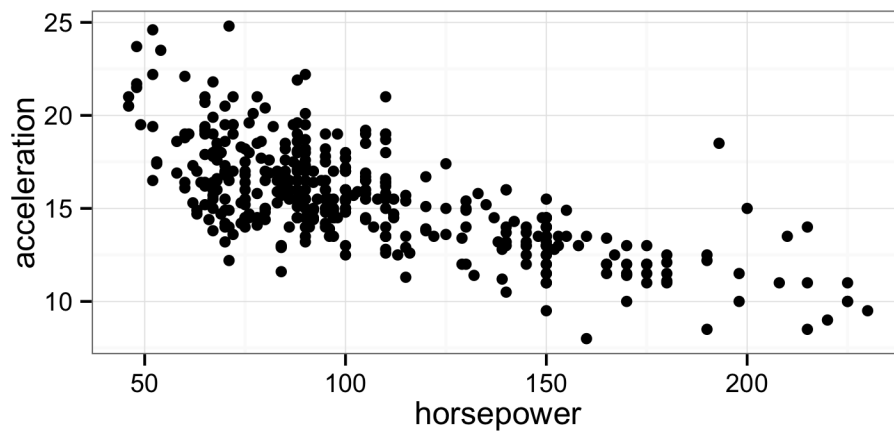


Figure 2: MPG vs number of cylinders

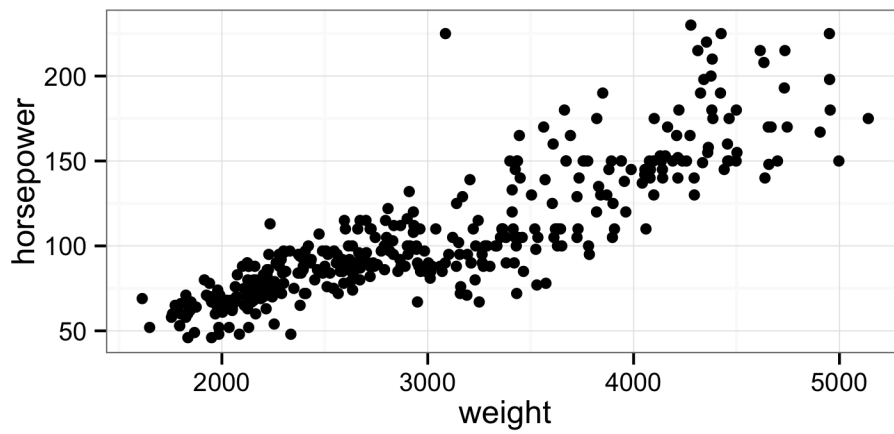
There isn't a particularly strong relationship between weight and acceleration (at least not one that is easily seen graphically).



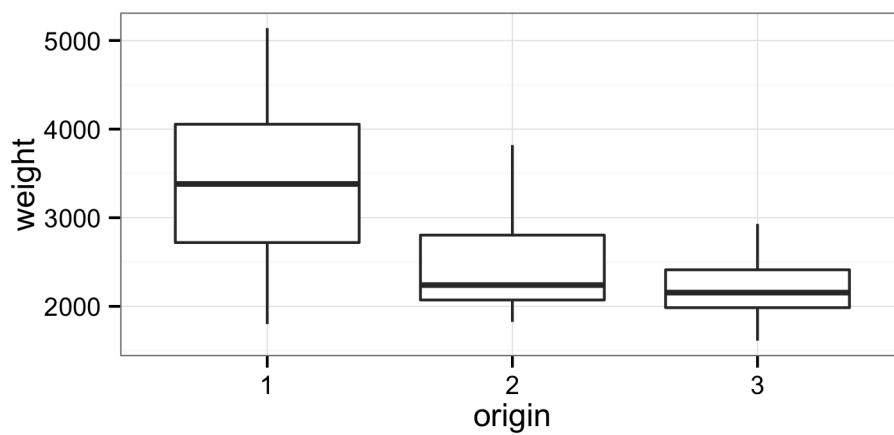
But there is a strong negative relationship between horsepower and acceleration.



And unsurprisingly, as the weight of a car increases, so does its horsepower.



Japanese cars (3) are usually lighter than American (1) and European (2) cars.



Japanese cars also have higher MPG.

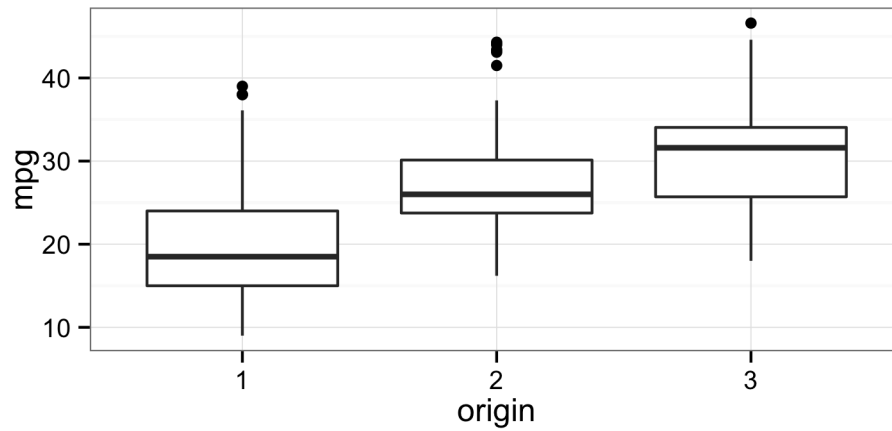


Figure 3: MPG by origin

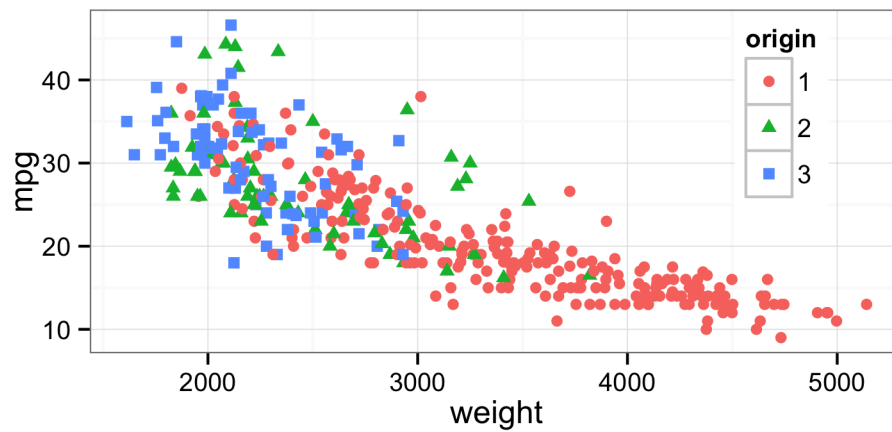


Figure 4: MPG by origin

Part f

As shown in **Part e**, the year (Figure 1), number of cylinders (Figure 2), origin (Figure 3), and weight (Figure 4) of a car are all useful in predicting MPG. Displacement is also a useful predictor, but acceleration is not (see Figures 5 and 6 below).

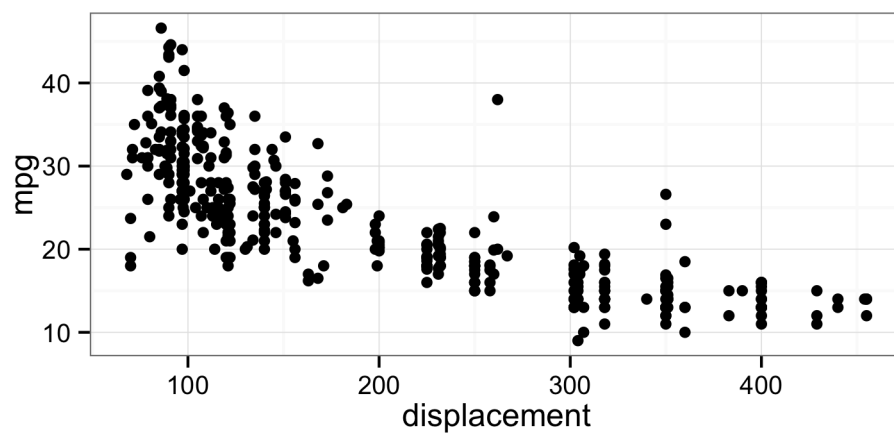


Figure 5: MPG vs displacemet

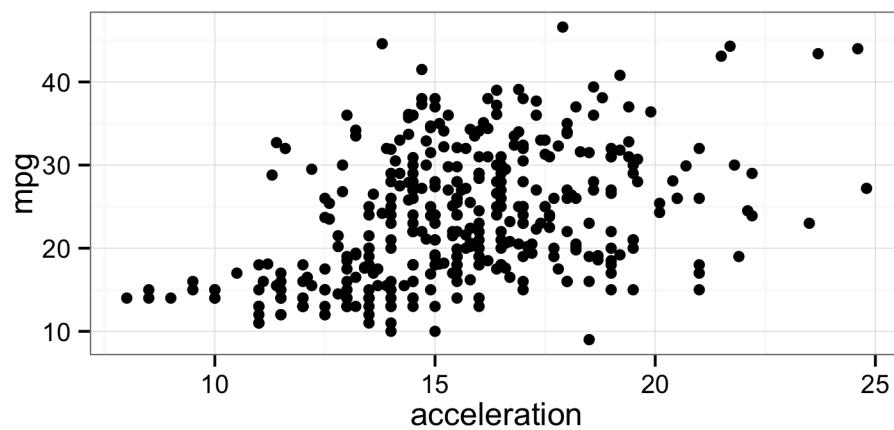


Figure 6: MPG vs acceleration