STA 032 Homework 1

Hardy Jones
999397426
Professor Melcon
Fall 2014

(a)

GPA		Semester		Year	
Min.	1.000	Fall	156	Min.	2007
1st Qu.	2.967	Spring	352	1st Qu.	2007
Median	3.250			Median	2007
Mean	3.191			Mean	2007
3rd Qu.	3.500			3rd Qu.	2008
Max.	4.000			Max.	2008

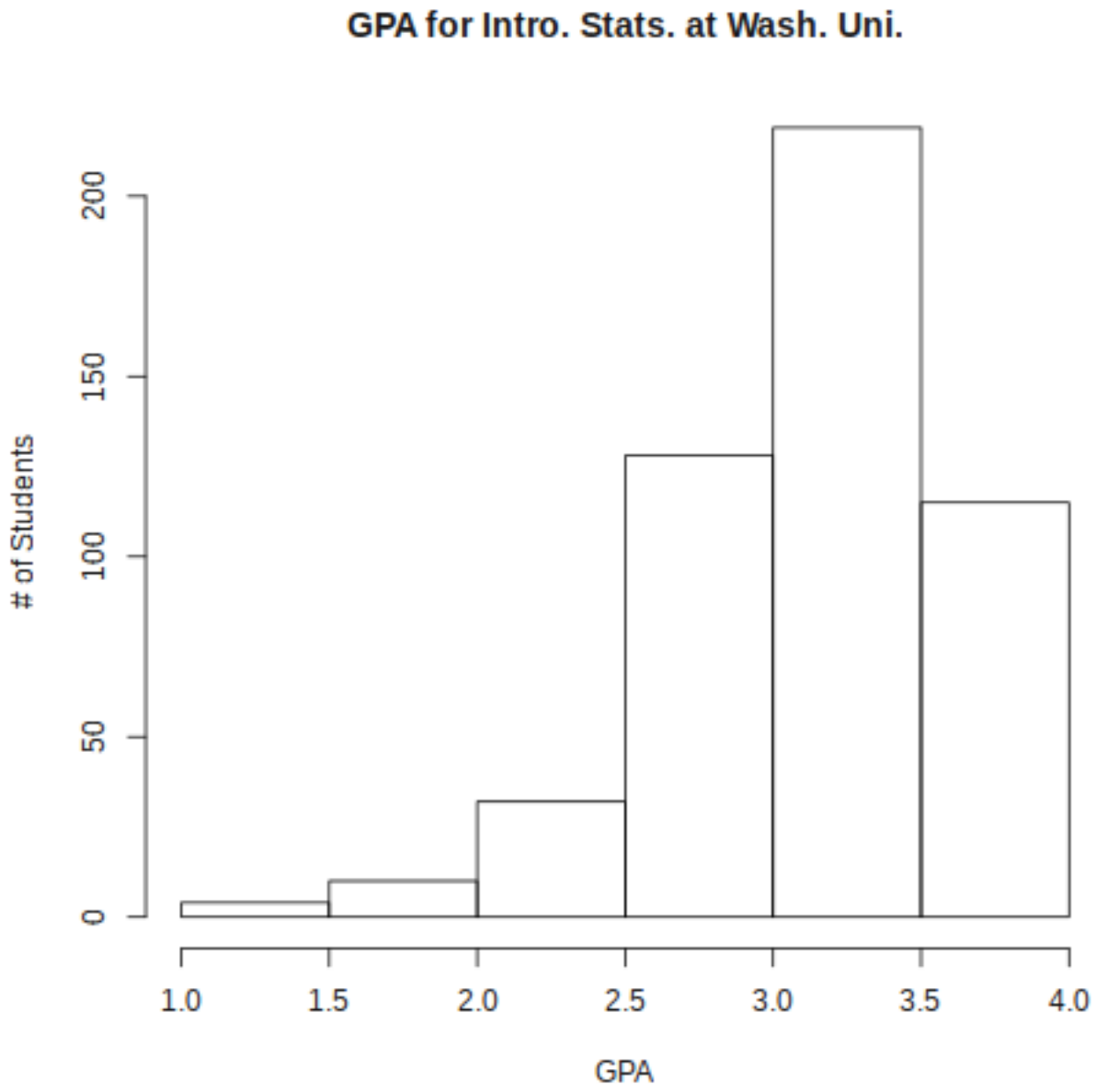
(b)

mean	standard deviation
3.190807	0.4716277

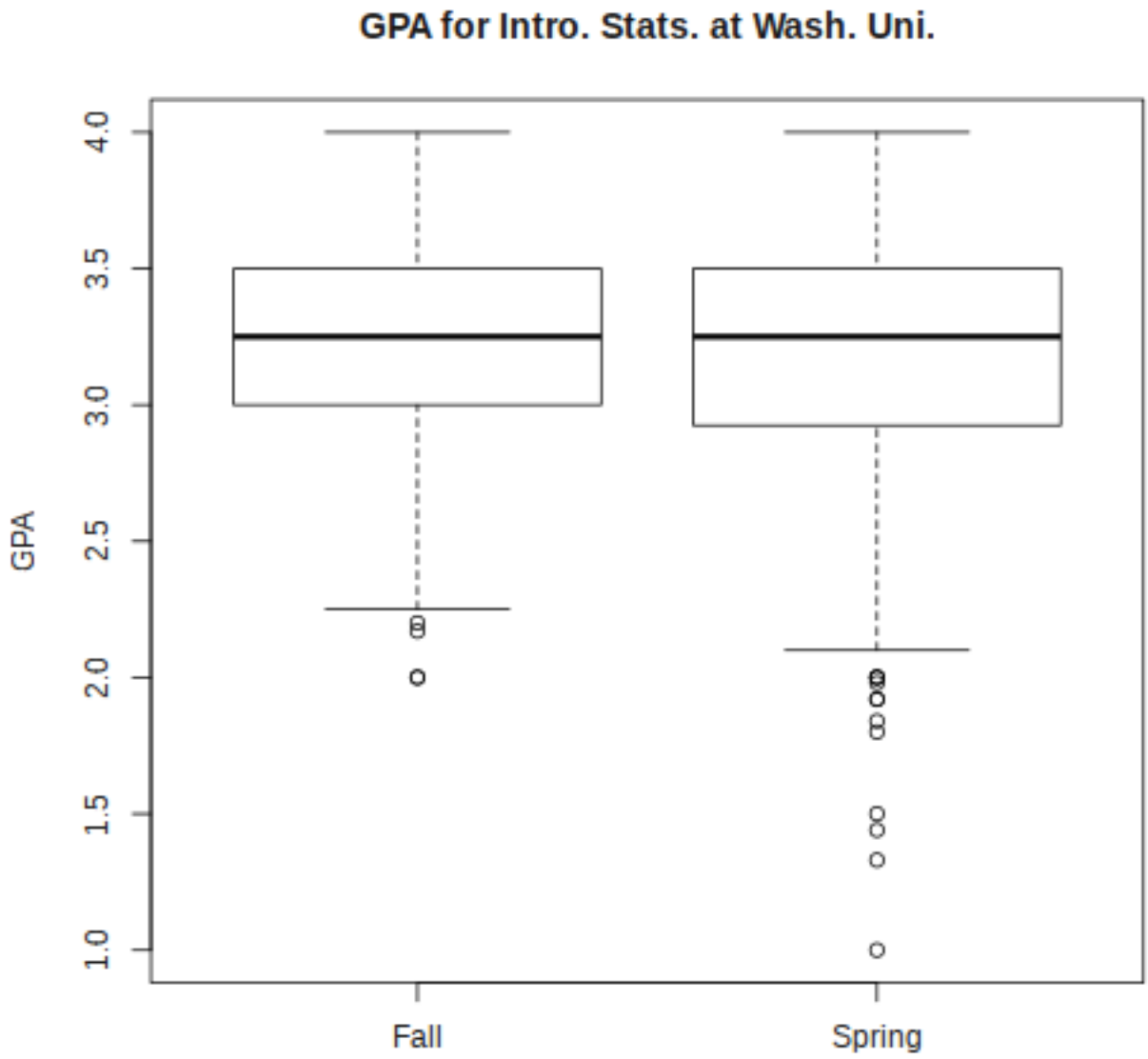
(c)

Fall	Spring
156	352

(d) There are 508 rows of data.


GPA for Intro. Stats. at Wash. Uni.
# of Students
0
50
100
150
200
1.0
1.5
2.0
2.5
3.0
3.5
4.0
GPA

(e)



(f)

(g)

	Q_1	Q_2	Q_3	90^{th} percentile
GPA	2.9675	3.2500	3.5000	3.7500

(h)

	5^{th} percentile	95^{th} percentile
GPA	2.4675	3.8725

(i)

	5^{th} percentile	95^{th} percentile
GPA	2.25	3.84

(j) Based on the data, it would appear that students have a higher GPA in the fall than they do in the spring. This result comes from the fact that the all of the quartiles in

fall are higher than they are in the spring, and that there are fewer outliers in the fall than there are in the spring.

However, since this is a sample of 508 students from one college over the course of four semesters, this is not indicative one way or the other of the population of students taking introductory statistics. More conclusive results could be found by including data from other schools and for more years.

Appendix A R code

We start by loading the data into R.

```
> data <- read.csv('IntroStatData.csv')
```

(a)

We can simply call the *summary* function to create the summary.

```
> summary(data)
```

(b)

We can get the mean and standard deviation with the functions *mean* and *sd*

```
> mean(data$GPA)
> sd(data$GPA)
```

(c)

The *table* function summarizes the data as categorical. In this case, it categorizes *data$Semester* into *Fall* and *Spring*

```
> table(data$Semester)
```

(d)

The *nrow* function counts the number of rows in the data set.

```
> nrow(data)
```

(e)

```
> png(filename="R_hw1/e.png")
> hist(data$GPA, main="GPA for Intro. Stats. at Wash. Uni.", xlab="GPA",
      ylab="# of Students")
> dev.off()
```

(f)

```
> fallGPA <- subset(data, Semester=="Fall", select=GPA)$GPA
> springGPA <- subset(data, Semester=="Spring", select=GPA)$GPA
> png(filename="R_hw1/f.png")
> boxplot(fallGPA, springGPA, main="GPA for Intro. Stats. at Wash. Uni.",
      names=c("Fall", "Spring"), ylab="GPA")
> dev.off()
```

(g)

We can use the *quantile* function to compute these values.

```
> quantile(data$GPA, c(0.25, 0.5, 0.75, 0.9))
```

(h)

We can use the *quantile* function to compute these values.

```
> quantile(fallGPA, c(0.05, 0.95))
```

(i)

We can use the *quantile* function to compute these values.

```
> quantile(springGPA, c(0.05, 0.95))
```