

STAT 5014 Homework 3

Samantha Sunshine

9/18/17

Problem 4

The programming style guides seem very useful for making R documents look nicer and more readable. There are different suggestions for the same types of formatting, which enables us to choose the style that we like the best. Specifically for me, I want to add more spacing and indentions between pieces of code to help make it more readable to other people.

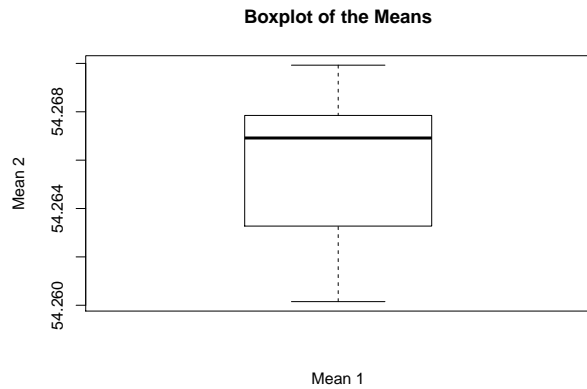
Problem 5

Lint showed me that I need to change the spacing I use around operators such as “<-” and “=”. I also should have a space after all commas and keep all variable and function names in lowercase letters.

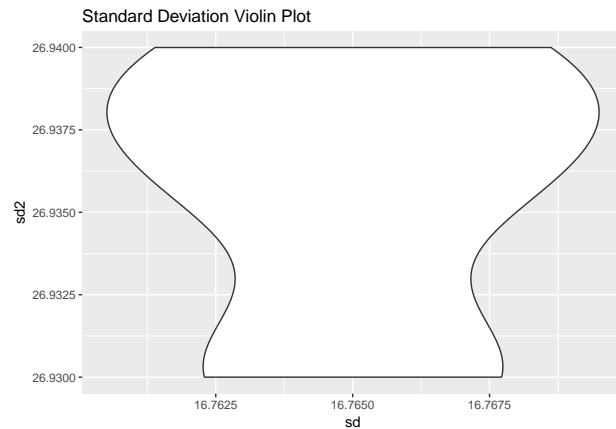
Problem 6

Table 1: Observers Data

Mean 1	Mean 2	Standard Deviation 1	Standard Deviation 2	Correlation Coefficient
54.26610	47.83472	16.76983	26.93974	-0.0655833
54.26873	47.83082	16.76924	26.93573	-0.0655833
54.26732	47.83772	16.76001	26.93004	-0.0655833
54.26327	47.83225	16.76514	26.93540	-0.0655833
54.26030	47.83983	16.76774	26.93019	-0.0655833
54.26144	47.83025	16.76590	26.93988	-0.0655833
54.26881	47.83545	16.76670	26.94000	-0.0655833
54.26785	47.83590	16.76676	26.93610	-0.0655833
54.26588	47.83150	16.76885	26.93861	-0.0655833
54.26734	47.83955	16.76896	26.93027	-0.0655833
54.26993	47.83699	16.76996	26.93768	-0.0655833
54.26692	47.83160	16.77000	26.93790	-0.0655833
54.26015	47.83972	16.76996	26.93000	-0.0655833



Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.
 Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.



Problem 7

Table 2: Blood Pressure Data

Day	Device 1	Device 2	Device 3	Doctor 1	Doctor 2	Doctor 3
1	133.34	133.36	133.45	126.54	127.36	131.88
2	110.94	110.85	110.92	124.69	128.86	132.39
3	118.54	118.56	118.67	125.46	129.43	134.43
4	137.94	137.80	137.77	125.95	130.72	134.28
5	139.52	139.62	139.59	125.90	130.13	134.44
6	139.23	139.11	139.36	127.85	132.03	137.37
7	117.96	117.81	117.85	125.55	132.05	132.17
8	119.59	119.42	119.48	125.80	129.87	134.97
9	116.12	116.00	115.93	125.11	128.09	133.97
10	128.38	128.48	128.41	125.75	131.94	132.68
11	125.17	125.25	125.34	128.77	130.05	134.75
12	134.62	134.41	134.55	125.26	131.13	134.29
13	136.14	136.07	136.22	126.26	130.91	133.38
14	131.21	131.03	130.96	125.68	128.83	135.67
15	132.51	132.86	132.65	124.47	129.46	134.39

First, I made “url” the name of the link to the blood pressure data.

```
url <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BloodPressure.dat"
```

Next, I named a table “bloodpressure_raw”, which included the data from url with no header and skipping the first two lines.

```
bloodpressure_raw <- read.table(url, header = F, skip = 2, fill = T, stringsAsFactors = F)
```

Then I named a new dataset “bloodpressure_tidy” which excluded column 5 from “bloodpressure_raw”.

```
bloodpressure_tidy <- bloodpressure_raw[, -5]
```

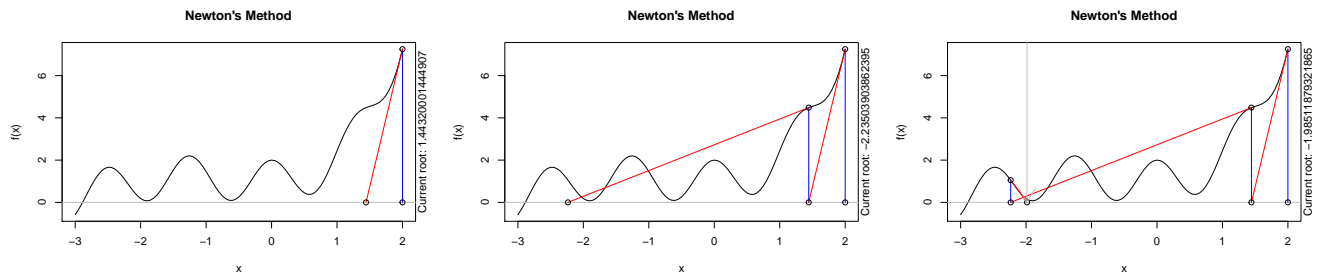
The last step was to rename the columns.

```
colnames(bloodpressure_tidy) <-  
  c("Day", "Device 1", "Device 2", "Device 3", "Doctor 1", "Doctor 2", "Doctor 3")
```

Table 3: Blood Pressure Data Summary

Day	Device 1	Device 2	Device 3	Doctor 1	Doctor 2	Doctor 3
Min. : 1.0	Min. :110.9	Min. :110.8	Min. :110.9	Min. :124.5	Min. :127.4	Min. :131.9
1st Qu.: 4.5	1st Qu.:119.1	1st Qu.:119.0	1st Qu.:119.1	1st Qu.:125.4	1st Qu.:129.1	1st Qu.:133.0
Median : 8.0	Median :131.2	Median :131.0	Median :131.0	Median :125.8	Median :130.1	Median :134.3
Mean : 8.0	Mean :128.1	Mean :128.0	Mean :128.1	Mean :125.9	Mean :130.1	Mean :134.1
3rd Qu.:11.5	3rd Qu.:135.4	3rd Qu.:135.2	3rd Qu.:135.4	3rd Qu.:126.1	3rd Qu.:131.0	3rd Qu.:134.6
Max. :15.0	Max. :139.5	Max. :139.6	Max. :139.6	Max. :128.8	Max. :132.1	Max. :137.4

Problem 8



Problem 9

Warning in model.response(mf, "numeric"): using type = "numeric" with a factor response will be ignored

Warning in Ops.factor(y, z\$residuals): '-' not meaningful for factors

Warning in model.response(mf, "numeric"): using type = "numeric" with a factor response will be ignored

Warning in Ops.factor(y, z\$residuals): '-' not meaningful for factors

Table 4: Car Data

Defect Code	License Plate	Make	Model
116	64BBT6	TOYOTA	TOYOTA COROLLA VERSO; 1.8 16V VVT-I MMT
118	12GRJF	VOLVO	S60; 2.4D AUT.
517	04VPPR	TOYOTA	TOYOTA VERSO-S
084	23NNVP	JAGUAR	JAGUAR XK8 COUPE
085	57BVKZ	VOLKSWAGEN	POLO
710	TLHZ93	RENAULT	ESPACE; GRAND 2.0 16V TURBO E4
135	88PLR3	CITROEN	BERLINGO; MULTISPACE 1.8I
521	4VZP09	PEUGEOT	308
130	83LFZK	TOYOTA	TOYOTA RAV4
860	51HGZ7	VOLKSWAGEN	PASSAT; VARIANT 110 KW
067	AB3003	OPEL	ZAFIRA-A; Z1.6XE
217	54JHN1	MERCEDES-BENZ	CLA 180
059	09DRTN	RENAULT	KANGOO
387	70RXKB	PEUGEOT	407
218	ZNJF34	PEUGEOT	206; CC 1.6 16V AUT.
417	OG92YN	SUZUKI	ALTO
333	13PSVB	OPEL	CORSA
407	BZHL94	SKODA	FABIA
112	83VPZ5	HYUNDAI	I40
362	59HZF7	RENAULT	LAGUNA; GRAND TOUR 1.8 16V 89KW S2005
178	67TPGX	AUDI	A3 SPORTBACK E-TRON
648	17PGVB	ALFA ROMEO	ALFA BRERA; 3.2 JTS Q4
864	HVSH75	VOLKSWAGEN	GOLF
313	83NHRX	VOLKSWAGEN	NEW BEETLE; 85 KW
170	XRPH69	BMW	3ER REIHE
608	82JGRR	AUDI	AUDI A6 ALLROAD QUATTRO
674	OH30DX	VOLKSWAGEN	GOLF
422	32TLXH	SAAB	SAAB 9-5; SEDAN 2.0 T
029	45JDPR	HYUNDAI	ATOS-PRIME; 1.1

Table 5: Car Summary Data

Defect Code	License Plate	Make	Model
Length:29	Length:29	Length:29	Length:29
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

There are 16 unique makes of cars.

Make
TOYOTA
VOLVO
JAGUAR
VOLKSWAGEN
RENAULT
CITROEN
PEUGEOT
OPEL

Make
MERCEDES-BENZ
SUZUKI
SKODA
HYUNDAI
AUDI
ALFA ROMEO
BMW
SAAB

There are no two defects that are reported twice in my data. I created a table of the 4 car makes with the most defects. One make had 5 defects, three makes had 3 defects, so I omitted the makes with 2 or 1 defects.

Table 7: Number of Defects in Makes

	x	freq
9	PEUGEOT	3
10	RENAULT	3
14	TOYOTA	3
15	VOLKSWAGEN	5

Below is a table of the correlation relationship in the number of defects by each make.

```
## Warning in Ops.factor(r, 2): '^' not meaningful for factors
```

Table 8: Fitting linear model: count(car_data[, 3])

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.795	NA	NA	NA
freq	1.492	NA	NA	NA

Below is the same table, but this time for each model.

```
## Warning in Ops.factor(r, 2): '^' not meaningful for factors
```

Table 9: Fitting linear model: count(car_data[, 4])

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.96	NA	NA	NA
freq	0.5185	NA	NA	NA