

# stat19000project01solutions

January 30, 2020

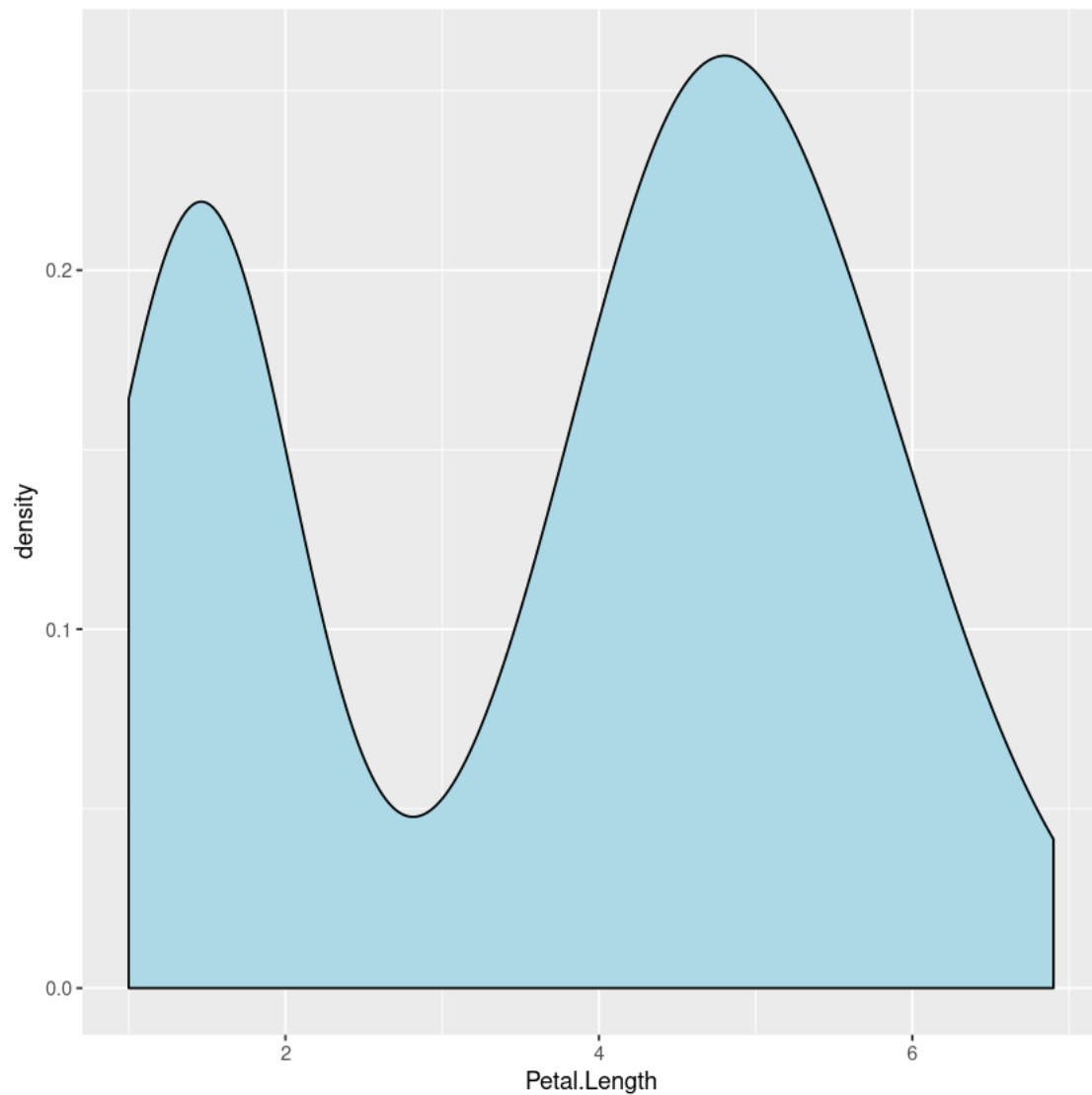
## 1 Project 1 Solutions

by Kevin Amstutz

```
[1]: # 1a
install.packages("ggplot2")
library("ggplot2")
ggplot(data=iris, aes(x=Petal.Length)) + geom_density(fill="lightblue")
```

Installing package into ‘/home/kamstut/R/scholar/3.6.1\_gcc-6.3.0\_obgeneric’  
(as ‘lib’ is unspecified)

Warning message in install.packages("ggplot2"):  
"installation of package ‘ggplot2’ had non-zero exit status"



```
[2]: # 1b  
      # the text is rendered
```

```
[3]: # 1c  
      # .ipynb, .r, .html, .md, .rst, .tex, .pdf
```

```
[4]: # 2a  
myDF <- read.csv('/class/datamine/data/8451/The_Complete_Journey_2_Master/  
↳5000_transactions.csv')
```

```
[5]: # 2b  
myresults <- split(myDF, myDF$STORE_R)
```

```
[7]: # 2c
class(myresults)
length(myresults)
names(myresults)
```

'list'

4

1. 'CENTRAL' 2. 'EAST ' 3. 'SOUTH ' 4. 'WEST '

```
[8]: # 2d
dim(myresults[["CENTRAL"]])
head(myresults[["CENTRAL"]])
```

1. 2463343 2. 9

A data.frame: 6 × 9

	BASKET_NUM <dbl>	HSHD_NUM <dbl>	PURCHASE_ <fct>	PRODUCT_NUM <dbl>	SPEND <dbl>	UN <dbl>
13	462	807	03-JAN-16	208846	3.99	1
15	591	999	03-JAN-16	93067	2.00	1
20	834	907	04-JAN-16	5423151	1.88	1
25	1424	4231	05-JAN-16	5180739	1.67	1
26	1494	1944	03-JAN-16	4667776	1.99	1
30	1583	3135	03-JAN-16	95584	1.00	1

```
[9]: # 2e
centralresults <- myDF[myDF$STORE_R == "CENTRAL", ]
dim(centralresults)
head(centralresults)
```

1. 2463343 2. 9

A data.frame: 6 × 9

	BASKET_NUM <dbl>	HSHD_NUM <dbl>	PURCHASE_ <fct>	PRODUCT_NUM <dbl>	SPEND <dbl>	UN <dbl>
13	462	807	03-JAN-16	208846	3.99	1
15	591	999	03-JAN-16	93067	2.00	1
20	834	907	04-JAN-16	5423151	1.88	1
25	1424	4231	05-JAN-16	5180739	1.67	1
26	1494	1944	03-JAN-16	4667776	1.99	1
30	1583	3135	03-JAN-16	95584	1.00	1

```
[10]: # 3a
loans <- read.csv("https://raw.githubusercontent.com/rfordatascience/
↳tidytuesday/master/data/2019/2019-11-26/loans.csv")
head(loans, n=3)
```

		agency_name <fct>	year <int>	quarter <int>	starting <dbl>	added <dbl>	total <dbl>
A data.frame: 3 × 10	1	Account Control Technology, Inc.	15	4	5807704381	1040570567	1226
	2	Allied Interstate, Inc.	15	4	3693337631	NA	1133
	3	CBE Group	15	4	2364391549	NA	8385

```
[11]: # 3b
sloans <- subset(loans, year %in% c(17,18) & total >= 1000000, c(1:2,
↪voluntary_payments:wage_garnishments))
```

```
[12]: # 3c
sloans[which(sloans$wage_garnishments == min(sloans$wage_garnishments)),]
sloans[which(sloans$voluntary_payments == min(sloans$voluntary_payments)),]
```

		agency_name <fct>	year <int>	voluntary_payments <dbl>	wage_garnishments <dbl>
A data.frame: 1 × 4	228	Pioneer Credit Recovery, Inc	18	238059.5	3907.52

		agency_name <fct>	year <int>	voluntary_payments <dbl>	wage_garnishments <dbl>
A data.frame: 1 × 4	223	ERS	18	75046.88	387079.8

```
[13]: # 3d
sloans$ratio_forced <- sloans$wage_garnishments/sloans$voluntary_payments
sloans[which.max(sloans$ratio_forced),]
sloans[which.min(sloans$ratio_forced),]
```

		agency_name <fct>	year <int>	voluntary_payments <dbl>	wage_ga <dbl>
A data.frame: 1 × 5	173	Financial Asset Management Systems, Inc.	17	200927.1	2916442

		agency_name <fct>	year <int>	voluntary_payments <dbl>	wage_garnishments <dbl>	ratio_forced <dbl>
A data.frame: 1 × 5	264	Pioneer	18	7098507	44393.48	0.006253918

```
[14]: # 3e
pie(c(sum(sloans$wage_garnishments), sum(sloans$voluntary_payments)),
↪labels=c("garnishments", "voluntary"), col=c("tomato", "lightblue"))
```

