Shefali Emmanuel's Data210
Homework #3
Spring 2020

Type neatly and show all your work!

This homework will let you use Rio and ggplot2 to create data visualization on iris.csv through command line. Please copy the command and insert the results and graph (.png) in this document.

1): Statistics and correlation between features(20pt)

(a) Type the command to calculate the mean, minimum, standard deviation of feature **sepal_width** and show the results.

```
[[/data/L07]$ <iris.csv Rio -e 'mean(df$sepal_width)'
[3.057333[/dat<iris.csv Rio -e 'min(df$sepal_width)'
[2[/data/L07]$ <iris.csv Rio -e 'sd(df$sepal_width)'
 0.4358663[/data/L07]$
```

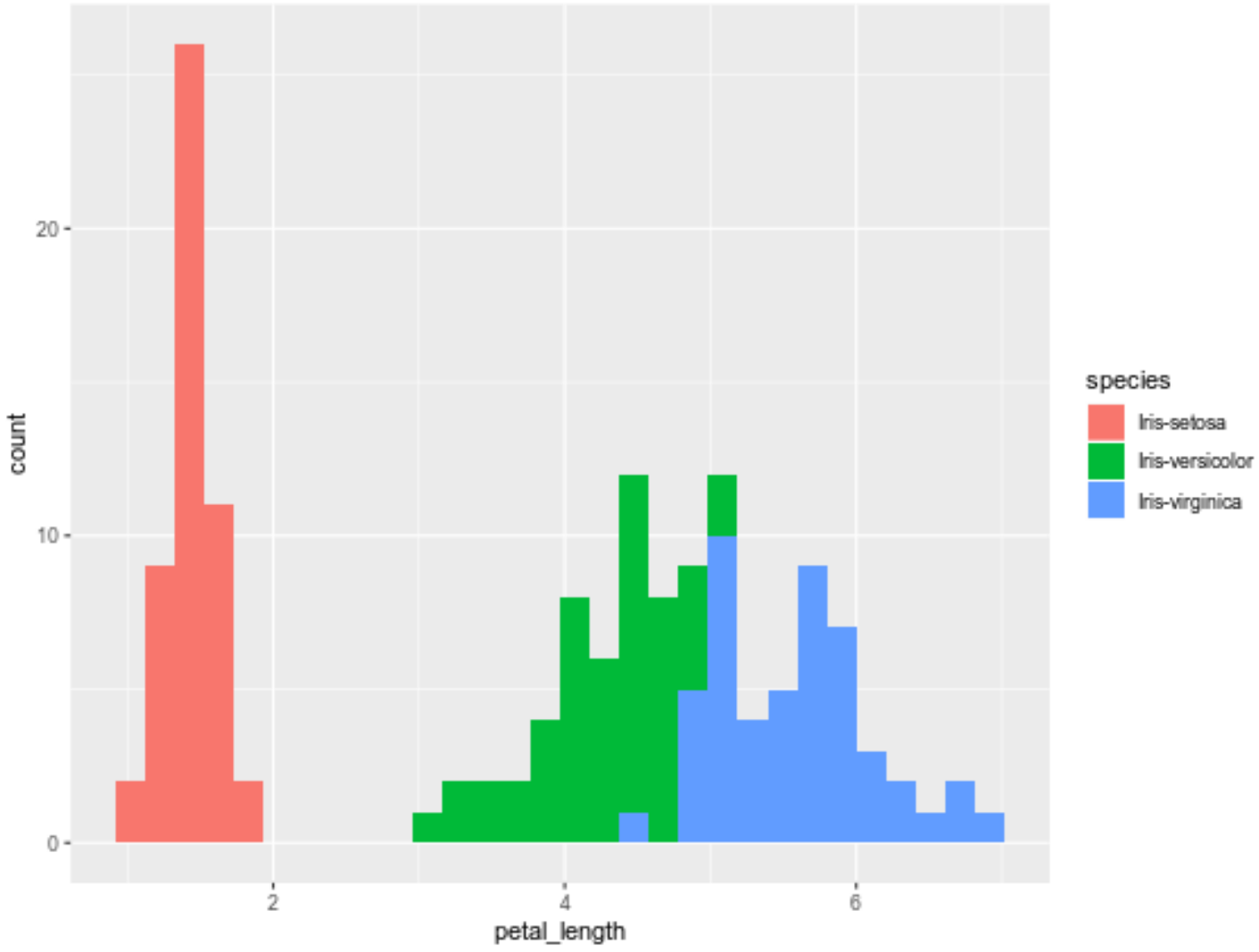
(b) Type the command to calculate the correlation between features **sepal_width** and **petal_width**, and show the results.

```
[[/data/L07]$  <iris.csv csvcut -c sepal_width,petal_width | Rio -f cor | csvlook
 | sepal_width | petal_width |
 | ----------- | ----------- |
 |       1.000… |      -0.366… |
 |      -0.366… |       1.000… |
 [/data/L07]$
```

2): Histogram

Type the command to plot histogram on feature **petal_length**, show the 3 species in different color, and inset the graph here.

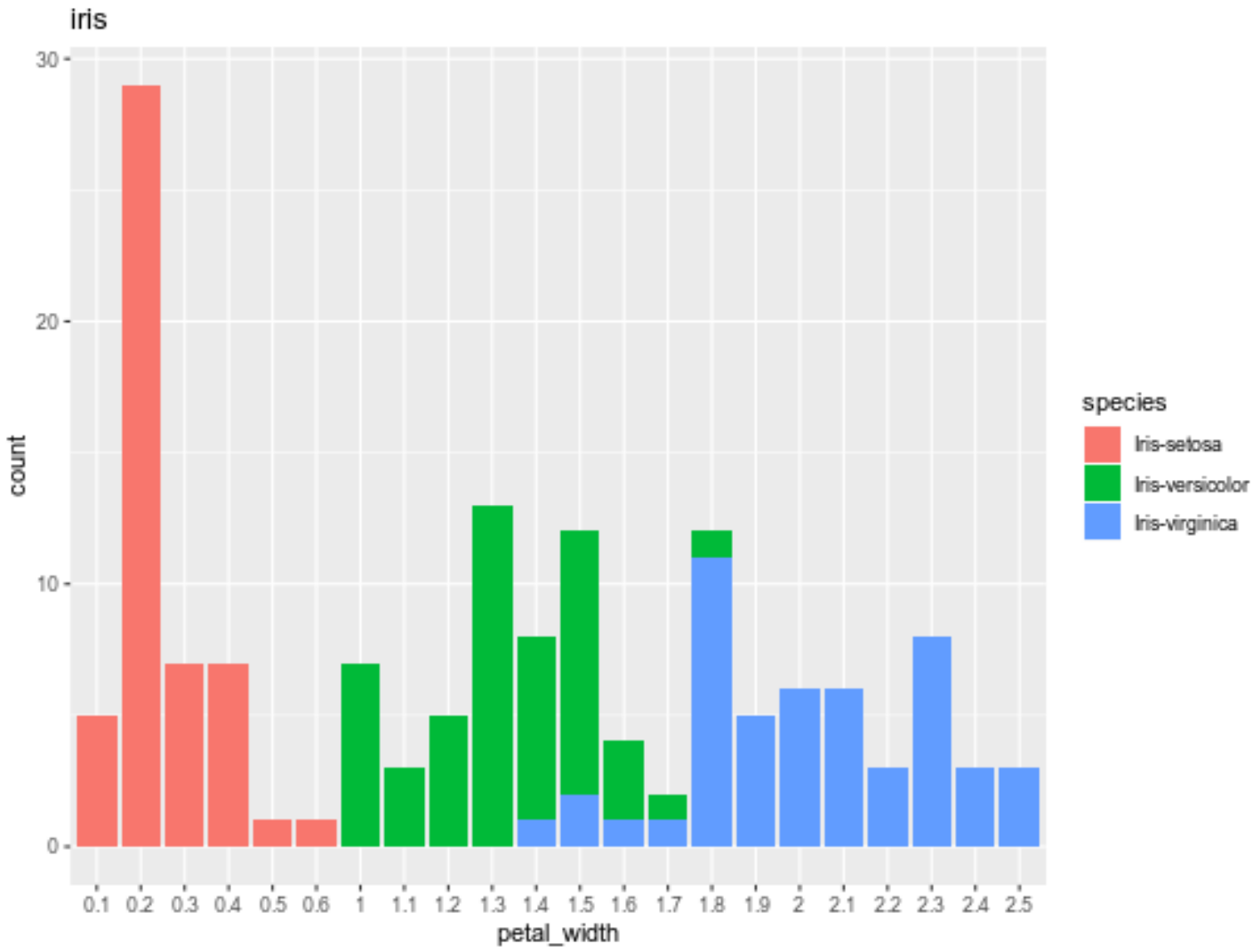
```
[[/data/L07]$ <iris.csv Rio -ge 'g+geom_histogram(aes(petal_length, fill=species))' > iris_petalLength_his.png
/usr/bin/Rio: line 128: warning: command substitution: ignored null byte in input
```



3): Bar plot

Type the command to plot bar graph on feature **petal_width**, show the 3 species in different color, change the xlabel as petal_width, and inset the graph here.

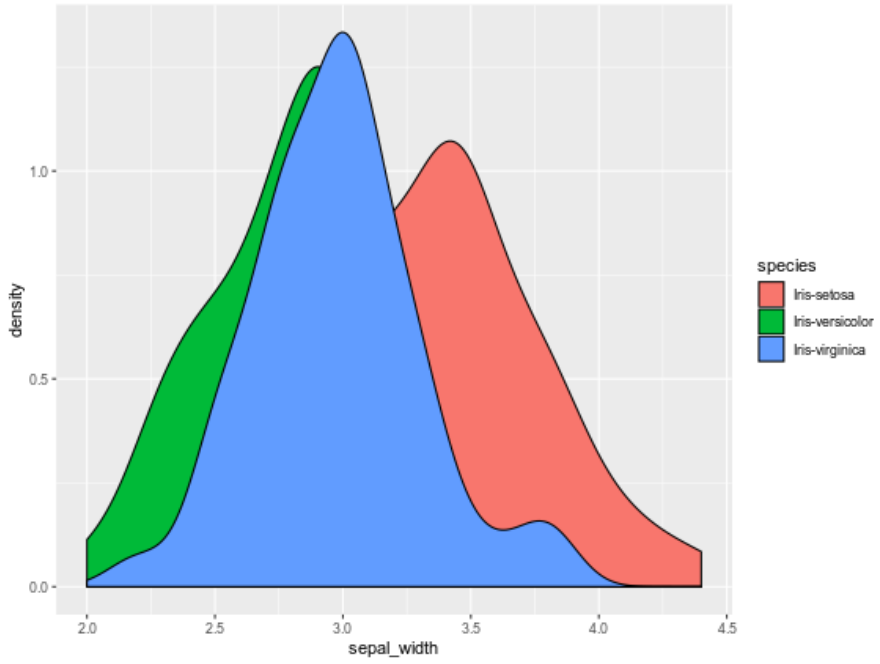
```
[/data/L07]$ <iris.csv Rio -ge 'g+geom_bar(aes(factor(petal_width),fill=species))+labs(x="petal_width",y="count",title="iris")' > iris_petalWidth_bar.png
/usr/bin/Rio: line 128: warning: command substitution: ignored null byte in input
```



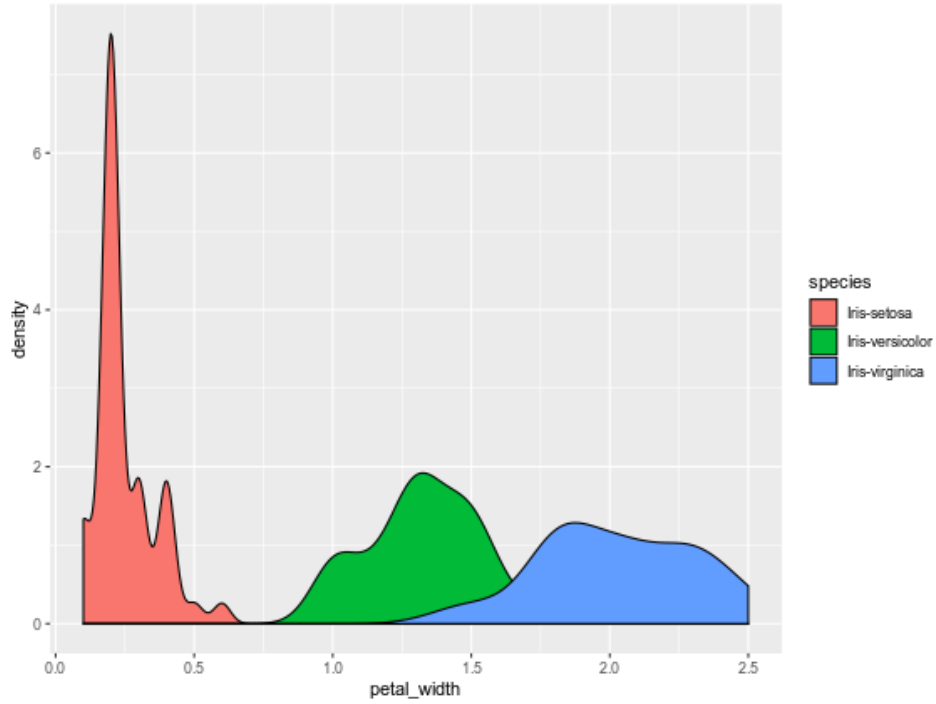
4): Density plot

Type the command to plot density on feature **sepal_width** and **petal_width**, each has 3 gaussian distribution, show the 3 species in different color, and inset the 2 plots here.

```
[/data/L07]$ <iris.csv Rio -ge 'g+geom_density(aes(sepal_width,fill=species))+xlab("sepal_width")' > iris_septalWidth_pdf.png
/usr/bin/Rio: line 128: warning: command substitution: ignored null byte in input
```



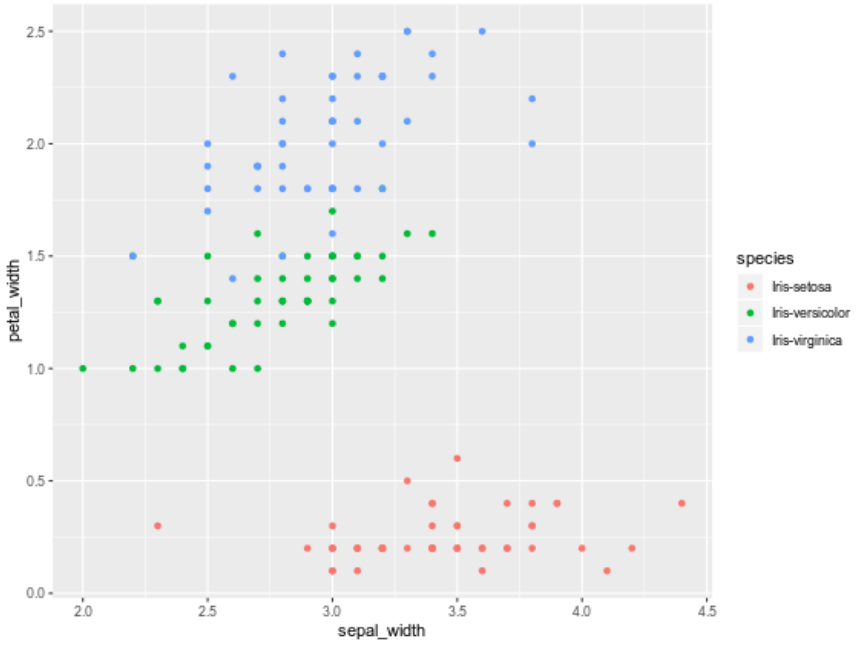

```
[/data/L07]$ <iris.csv Rio -ge 'g+geom_density(aes(petal_width,fill=species))+xlab("petal_width")' > iris_petalWidth_pdf.png
/usr/bin/Rio: line 128: warning: command substitution: ignored null byte in input
```



5): Scatter plot

Type the command to draw scatter plot on feature **sepal_width** vs **petal_width**, and **sepal_length** vs **pedal_length**, showing the 3 species in different colors, and insert the 2 plots here.

```
[/data/L07]$  <iris.csv Rio -ge 'g +geom_point(aes(sepal_width, petal_width, color=species))' > iris_septalWidth_petalWidth.png
/usr/bin/Rio: line 128: warning: command substitution: ignored null byte in input
```




```
[/data/L07]$  <iris.csv Rio -ge 'g +geom_point(aes(sepal_length, petal_length, color=species))' > iris_septalLength_petalLength.png
/usr/bin/Rio: line 128: warning: command substitution: ignored null byte in input
```