HW 10 Instructional Project 1

Dev Baiju

Due: Nov 15, 2024 9:30 AM

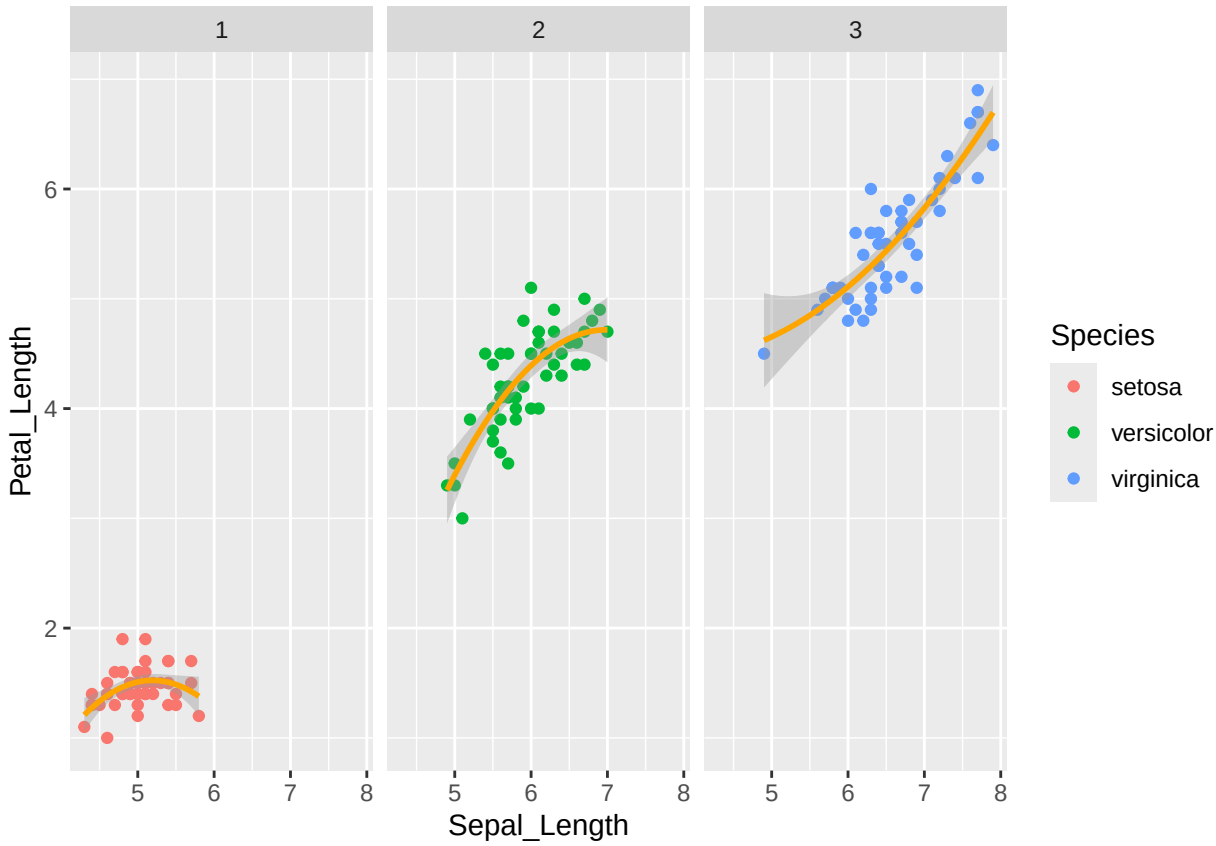
Problems

1. ggplot2 (prepared by Connor Curtiss, Pradeep Maripala, Ariadna Orbe Vivero, Behrooz Khalil Loo, Bowen Su).

Complete the two questions on the slides: ggplot2.pptx, on ICON. Answer for the first question:

```
library(ggplot2)
attach(iris)
df <- data.frame(Sepal_Length = Sepal.Length, Petal_Length = Petal.Length)
ggplot(df, aes(x = Sepal_Length, y = Petal_Length))+
  geom_point(aes(color = Species))+
  geom_smooth(span = 20, color = "orange")+
  facet_wrap(~ as.numeric(Species))
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Answer for the second question:

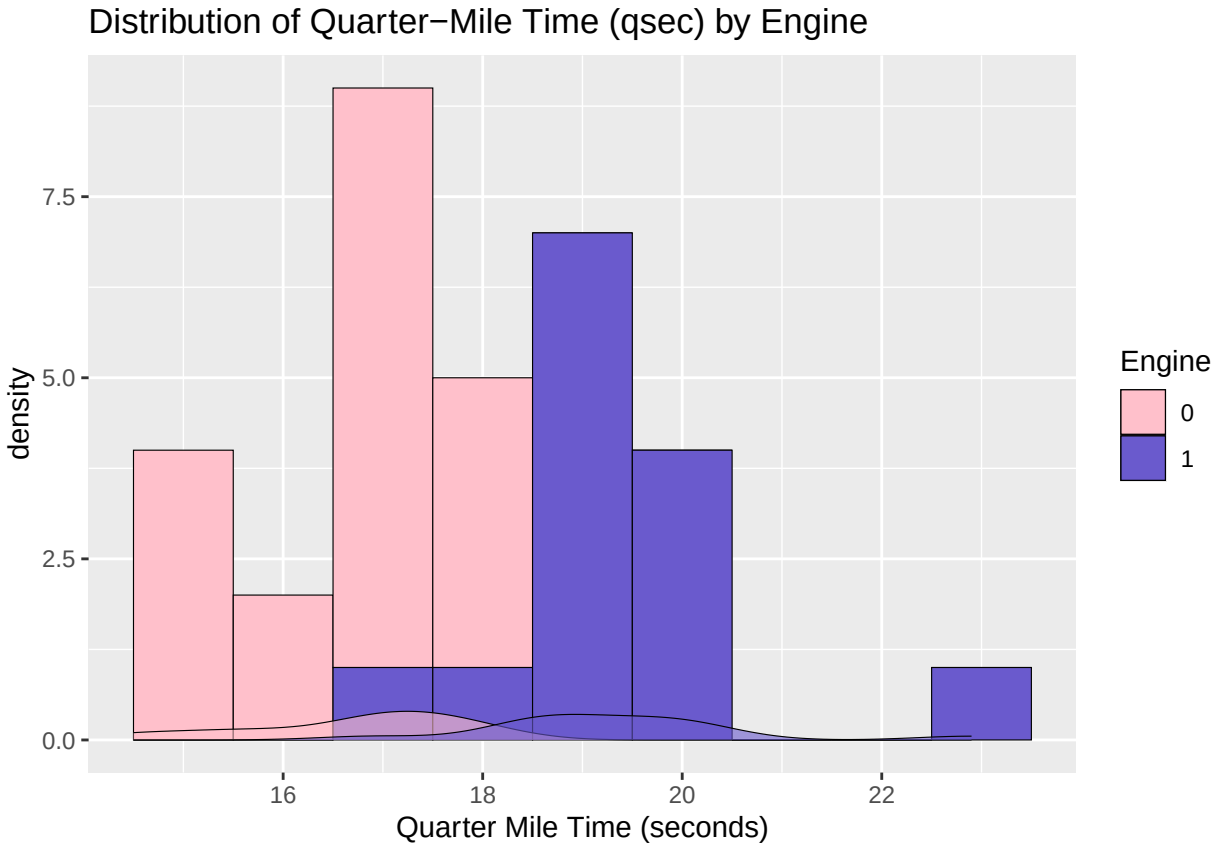
```
attach(mtcars)
```

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

```
df2 <- data.frame(Qmile = qsec, Engine = as.factor(vs))

ggplot(df2, aes(x = Qmile, fill = Engine))+
  geom_histogram(binwidth = 1, color = "black", size = 0.2)+
  scale_fill_manual(values = c("pink", "slateblue"))+
  geom_density(alpha = 0.6, size = 0.2)+
  ggtitle("Distribution of Quarter-Mile Time (qsec) by Engine")+
  xlab("Quarter Mile Time (seconds)")
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



2. Numpy and Pandas (prepared by Akhilesh Karra, Vaishnavi Soni, Nathan Munshower, Michael Sullivan, Dongwei Zhang)

Complete the questions on the slides: Numpy and Pandas.pptx, on ICON.

Answer to the first question:

```
import numpy as np

np.random.seed(42)
mat  = np.random.randint(1, 101, size=(10, 10))
print("The 10x10 matrix is\n", mat, "\n")
sliced_mat = mat[5:, 5:]
flatten_mat = sliced_mat.flatten()
print("Flattened matrix array: \n", flatten_mat, "\n")
newmat = flatten_mat + 5
print("After adding 5 to the flettened array: \n", newmat, "\n")
print("The median is ", np.median(newmat), "\n")
```

Answer to the second question:

```
import pandas as pd
uitem = pd.read_csv("/Users/devanandabaiju/Desktop/python/u.item", sep='|',
                    encoding='Latin-1', header=None)
print(uitem.head())
```

```
uitem.drop(uitem.columns[[3,4]], axis=1, inplace=True)
uitem['Sum']=sum(uitem.iloc[:,3:22])
print("\nSummary of the new sum column:")
print(uitem['Sum'].describe())
```

3. Practice of Julia (prepared by Phoebe Low and Ting-Hung Yu, STAT 5400 Fall 2020 alumni)

Watch the following videos and read Julia slides.pdf on the ICON site.

https://uicapture.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=f4b86ede-d182-4e87-9047-ac4800feeacf

https://uicapture.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=fd3ad007-deb3-4aa3-9f49-ac4800fee7b9

Finish the homework question assigned on the slides.

Use the online Julia platform https://repl.it/languages/julia or install Julia locally to write a function, say m(x), that finds mean and variance (no packages needed) simultaneously.

- Test on sequence of integers 1:10000.
- Paste the code below on the RMarkdown file, and attach a screenshot of results to the PDF.

```
function m(x)
  n = size(x)[1]
  meanx = 0
  varx = 0
  for i in 1:n
    meanx += x[i]
  end
  meanx = meanx/n
  for i in 1:n
    varx += (x[i] - meanx)^2
  end
  varx = varx/n
  return (mean = meanx, variance = varx)
end
println(m(collect(1:10000)))
```

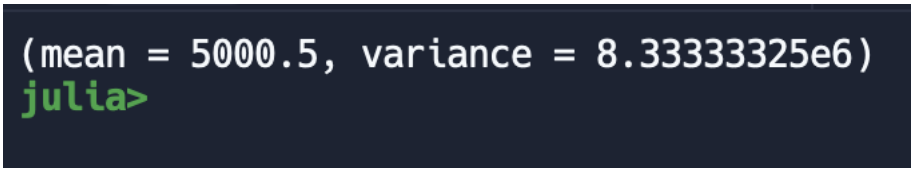


Figure 1: Julia Output

4. Connect R to ChatGPT Watch the following videos, and explore using ChatGPT in R throught API. You do not need to submit anything for this question. https://www.youtube.com/watch?v=szPIuzQ-jco&ab_channel=AnalyticoHub