Part 1 in R

```
# specifying the path
path <- "/Users/music/Desktop/avocados.csv"
# reading contents of csv file
avocados <- read.csv(path)
# contents of the csv file
print (avocados)

# load libraries
library(dplyr)
library(rcompanion)
library(car)
library(ggplot2)
library(IDPmisc)

# Part 1 in R: Does the average price of avocados differ between Albany, Houston, and Seattle?

plotNormalHistogram(avocados$AveragePrice)

# SQRT, data is slightly positively screwed

avocados$AveragePriceSQRT = sqrt(avocados$AveragePrice)

plotNormalHistogram(avocados$AveragePriceSQRT)

# Bartlett & Fligner tests of homogeneity of variances

bartlett.test(AveragePriceSQRT ~ region, data=avocados)

fligner.test(AveragePriceSQRT ~ region, data=avocados)

# violated the assumption of homogeneity of variance.

# Welch's One-Way Test
ANOVA <- lm(AveragePriceSQRT ~ region, data=avocados)
Anova(ANOVA, Type="II", white.adjust=TRUE)

# significant at p < 2.2e-16, therefore, significant difference in average price per region aka
Albany, Houston, and Seattle

# Post Hocs with no adjustment
pairwise.t.test(avocados$AveragePriceSQRT, avocados$region, p.adjust="none")
# with adjustment
```

```
pairwise.t.test(avocados$AveragePriceSQRT, avocados$region, p.adjust="bonferroni")
# violation
pairwise.t.test(avocados$AveragePriceSQRT, avocados$region, p.adjust="bonferroni", pool.sd
= FALSE)

# Houston vs Albany was  < 2e-16
# [ reached getOption("max.print") -- omitted 35 rows ]


avocadosMeans <- avocados %>% group_by(region) %>% summarize(Mean =
mean(AveragePriceSQRT))
# Albany 1.243726, Houston 1.013421, and Seattle 1.186291
# The means differ slightly.
```

Part 2 in Python