

Student	Yan Yinan	62
---------	-----------	----

Part 1

	Available	Marks
Generate WineSamples dataframe	5	2
Add proline categories	10	6
Add ratio of hue to colour intensity	5	3
Generate WineSummary	5	5
Add descriptive data for alcohol and phenols	10	4
Identify high/v. high proline levels	10	4
Commenting and use of white space	5	5

Comments

Line 12 works, but the bit with the library doesn't. Remember to always check your filepaths before sharing a script.

```
> # load the file-"wineSamples.xlsx" into data frame wineData.
>
> wineData<- read.xlsx("C:/Users/Allen/Desktop/wineSamples.xlsx")
Error in read.xlsx.default("C:/Users/Allen/Desktop/wineSamples.xlsx") :
  File does not exist.
```

Good explanation for removing the NA data.

You've created an appropriate categorisation range for proline, but it would be far better to calculate the category limits in the script rather than hard coding them as you have.

Calculation of ratio works but is not particularly efficient.

Good generation of WineSummary.

Use of tapply() is good, but remember that it reflects the results so a 3x4 data structure returns as a 4x3. This is why the data won't bind to the summary dataframe.

```
> wineSummary<-cbind(wineSummary,
+                     mean_Alcohol_wineData,
+                     sd_Alcohol_wineData,
+                     mean_Total.Phenols_wineData,
+                     sd_Total.Phenols_wineData)
Error in data.frame(..., check.names = FALSE) :
  arguments imply differing number of rows: 3, 4
> # Display wineSummary
```

You could have used aggregate() instead.

```
> wineSummary
  Type number percentage_high_proline percentage_vhigh_proline
1 type 1      59                    0                        0
2 type 2      71                    0                        0
3 type 3      48                    0                        0
>
```

Your filtering doesn't work, which is why the percentages aren't calculated.

Commenting and use of white space is good, but it would have been good to use some programmer-defined functions.

Part 2

	Available	Marks
Create the countryData dataframe	5	2
Create the CovidSymmary dataframe	15	9
Check the data for normality and process accordingly	10	6
Perform the appropriate t-tests	5	5
Perform the appropriate ANOVA tests	10	6
Commenting and use of white space	5	5
Comments Line 13 works, but the bit with the library doesn't. Remember to always check your filepaths before sharing a script. <pre>> library(openxlsx) > # load the file-"CovidData.xlsx" into data frame countryData. > > countryData<- read.xlsx("C:/Users/Allen/Desktop/CovidData.xlsx") Error in read.xlsx.default("C:/Users/Allen/Desktop/CovidData.xlsx") : File does not exist.</pre> You haven't picked up the data for weeks 25 or 36. Otherwise the process works, although it's not particularly efficient. Good use of the Descriptive_statistics function. Good process for handling outliers, although it could be automated rather than repeated. The t-test is fine. The ANOVA process is fine, although you only needed to do it for new cases. You've also missed the post-hoc test for new cases but included it for all of the others. Commenting and white space are good. It would have been nice to see more code moved into functions.		