

Student	Wu Pengchun	56
---------	-------------	----

Part 1

	Available	Marks
Generate WineSamples dataframe	5	1
Add proline categories	10	4
Add ratio of hue to colour intensity	5	4
Generate WineSummary	5	3
Add descriptive data for alcohol and phenols	10	6
Identify high/v. high proline levels	10	5
Commenting and use of white space	5	2

Comments

You MUST put your name and student ID in a comment at the start of the script.

The comments must also explain if any extra packages are used, rather than just trying to load them.

```
> library(xlsx)
```

```
Error: package or namespace load failed for 'xlsx':
.onLoad failed in loadNamespace() for 'rJava', details:
call: fun(libname, pkgname)
error: JAVA_HOME cannot be determined from the Registry
```

In addition: warning message:

```
package 'xlsx' was built under R version 4.0.5
```

```
> result <- read.xlsx("E:\\R Programming\\Task 2\\wineSamples.xlsx",sheetIndex = 1)
Error in read.xlsx("E:\\R Programming\\Task 2\\wineSamples.xlsx", sheetIndex = 1) :
unused argument (sheetIndex = 1)
```

The data file should be in the same folder as the script.

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.

```
+ else if(i%/%400 == 4)
+   ProlineLevel = c(ProlineLevel,'v.high')
+ }
```

```
Error in if (i%/%400 == 0) ProlineLevel = c(ProlineLevel, "v.low") else if (i%/%400 == :
missing value where TRUE/FALSE needed
```

```
> ProlineLevel = factor(ProlineLevel)
```

```
> wineData$Proline.level = ProlineLevel
```

```
Error in `<-`data.frame(`*tmp*`, Proline.level, value = c(3L, 3L, 3L, :
replacement has 178 rows, data has 181
```

You have calculated the proline level correctly but have not added it to the dataframe.

Calculation of high/v.high is good, but it would be better to format the result as 46.2% as opposed to 0.462.

The calculations for the wine types are correct buy you have an extra row.

	Wine.Type	type.Num	Mean.Alcohol	Std.Alcohol	Mean.Phenols	Std.Phenols
1		59	13.74475	0.4621254	2.840169	0.3389614
2	Type 1	71	12.27873	0.5379642	2.258873	0.5453611
3	Type 2	48	13.15375	0.5302413	1.678750	0.3569709
4	Type 3	0	NA	NA	NA	NA

It's much more efficient to calculate the mean and standard deviation using the `aggregate()` function.
The `high/v.high` proline calculation fails due to the problems with the dataframe.

Part 2

	Available	Marks
Create the <code>countryData</code> dataframe	5	1
Create the <code>CovidSummary</code> dataframe	15	11
Check the data for normality and process accordingly	10	6
Perform the appropriate t-tests	5	3
Perform the appropriate ANOVA tests	10	6
Commenting and use of white space	5	4
Comments You MUST put your name and student ID in a comment at the start of the script. The comments must also explain if any extra packages are used, rather than just trying to load them. The data file should be in the same folder as the script. <pre>> result <- read.xlsx("E:\\R Programming\\Task 2\\CovidData.xlsx", sheetIndex = 1) Error in read.xlsx("E:\\R Programming\\Task 2\\CovidData.xlsx", sheetIndex = 1) : could not find function "read.xlsx"</pre> Good use of <code>which()</code> to build <code>CovidSummary</code>. Good display of the range data. Process for dealing with outliers is fine. The code for the t-test is fine but you must define the null and alternate hypotheses in the comments before performing the test. The ANOVA process is fine, but again you should define the null and alternative hypotheses before performing the test. Use of white space is good but the commenting could be better. There are a lot of extra random characters within it.		