Solution to exercise 2

Data exploration

Explore the `airquality` dataset

i) Load the `airquality` dataset

```
data(airquality)
```

```
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

ii) Assign a new name to the dataset (e.g. `airquality_1`)

```
airquality_1 <- airquality
```

iii) Get an overview of the dataset:
How many rows does the dataset have?

```
dim(airquality_1) # returns the number of rows and columns
```

```
## [1] 153   6
```

```
nrow(airquality_1) # returns the number or rows
```

```
## [1] 153
```

How many columns does the dataset have?

```
dim(airquality_1) # returns the number of rows and columns
```

```
## [1] 153   6
```

```
ncol(airquality_1) # returns the number or columns
```

```
## [1] 6
```

What class do the columns have? Can you guess?

```
class(airquality_1$Ozone)
```

```
## [1] "integer"
```

```
class(airquality_1$Solar.R)
```

```
## [1] "integer"
```

```
class(airquality_1$Wind)
```

```
## [1] "numeric"
```

iv) Calculate the mean temperature

```
# select column Temp
temp <- airquality_1$Temp # select with name
temp <- airquality_1[, 4] # select with number

# calculate the mean
mean(temp)
```

```
## [1] 77.88235
```

```
# or combined in one line
mean(airquality_1$Temp)
```

```
## [1] 77.88235
```

v) What is the maximum ozone content?

```
# select column Ozone
ozone <- airquality_1$Ozone
ozone <- airquality_1[, 1]

# calculate the maximum
max(ozone) # result is NA because ozone contains NA's
```

```
## [1] NA
```

```
max(ozone, na.rm = NA)
```

```
## [1] 168
```

```
# or combined in one line
max(airquality_1$Ozone, na.rm = TRUE)
```

```
## [1] 168
```

vi) In which month and day did the maximum ozone content occur?

```
# select column Ozone
ozone <- airquality_1$Ozone

# option 1
which.max(ozone) # row number of maximum
```

```
## [1] 117
```

```
airquality_1[117, ] # select the row with the maximum ozone measurement
```

```
##     Ozone Solar.R Wind Temp Month Day
## 117   168     238  3.4   81     8  25
```

```
# option 2
airquality_1[which.max(ozone), ]
```

```
##     Ozone Solar.R Wind Temp Month Day
## 117   168     238  3.4   81     8  25
```

```
# option 3
# select the row with the maximum ozone measurement and the
# Month and Day column
airquality_1[117, 5:6]
```

```
##     Month Day
## 117     8  25
```

vii) Calculate the mean temperature for the month May

```
# subset to the month May
airquality_may <- subset(x = airquality_1, Month == 5)
unique(airquality_may$Month) # check the result
```

```
## [1] 5
```

```
# select the column Temp
temp_may <- airquality_may$Temp

# calculate the mean
mean(temp_may)
```

```
## [1] 65.54839
```