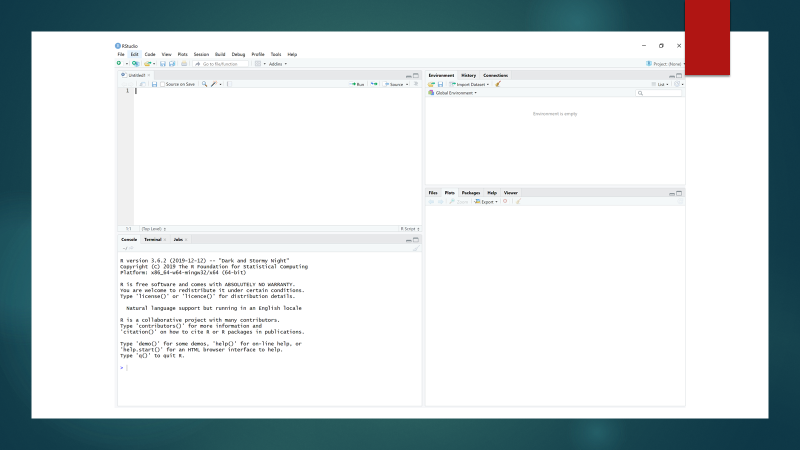
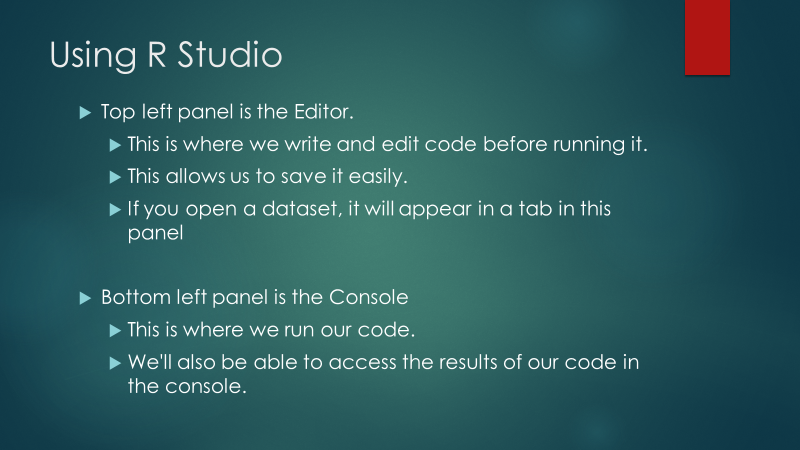
Learning Outcomes:

1. Set a working directory
2. Create a new R Script
3. Input data manually
4. Read in an existing dataset
5. Basic Descriptive statistics







**1) Set a working directory**

Create a folder on your personal drive for this module's labs.

Open up R Studio and set the working directory to your module folder using Session>Set Working Directory. When working on SCSS computers, make sure your working directory and the location where you save your R script is in your own student file location, not on the machine you are working on. Otherwise there will be an error as you do not have the appropriate user permissions.

Note that the CONSOLE window shows how you could set the working directory using code.

**2) Create a new R Script**

Use File>New File>R Script to create a new R Script

Use File>Save As to save the R Script as Lab1\_FirstnameLastname eg Lab1\_SusanConnolly

**3) Input data manually**

Type the following in the R Script Window and then click "run".

employee <- c('John Doe','Peter Gynn','Jolie Hope')

salary <- c(21000, 23400, 26800)

startdate <- as.Date(c('2010-11-1','2008-3-25','2007-3-14'))

Now you have three different vectors in your workspace:

- A character vector called employee, containing the names

- A numeric vector called salary, containing the yearly salaries

- A date vector called startdate, containing the dates on which the contracts started

Combine the three vectors into a data frame using the following code:

employ.data <- data.frame(employee, salary, startdate)

Note that you can highlight individual lines or sections of code before clicking "run" rather than running all the code in the window at once.

Type str(employ.data) to see the structure of the data frame.

In the environment window, click the blue arrow beside employ.data to see details of the data frame

In the environment window, click employ.data itself to open a view of the data itself in a tab beside your R Script

Type and run the following to change the column names

colnames(employ.data) = c("a","b","c")

Type and run the following code to change them back

colnames(employ.data) = c("employee","salary","startdate")

**4) Read in an existing dataset**

Download the csv (comma separated values) file Lab1.csv from Blackboard into your R working directory.

In the environment window, click Import Dataset and select From Text(readr)

Browse to the location of Lab1.csv and click "Import"

In the environment window, click the blue arrow beside Lab1 to see details of the data frame

In the environment window, click Lab1 itself to open a view of the data itself in a tab beside your R Script

If you get an error saying a package is missing, eg package “pillar”, try the following.

* Go to the packages tab in the lower right pane
* Click Install
* Type the name of the package r has asked for and click Install
* If a pop up appears, click No
* When the package has installed, find it in the list of packages and tick the checkbox
* Try to import the dataset again

Alternatively, you can read in the dataset using code.

In the R script window, type and run the following two lines of code. What is the difference between the two data frames Lab1\_true and Lab1\_false?

Lab1<-read.csv(file="Lab1.csv", header=TRUE)

Lab1\_false<-read.csv(file="Lab1.csv", header=FALSE)

If you get an error saying something like file or directory not found, check the following:

* Did you set your working directory?
* Is the csv file saved in the working directory?
* Is the csv file named Lab1.csv? (not Lab1 (1).csv or similar

**5) Basic Descriptive Statistics / Data Inspection**

We have 6 columns corresponding to the 6 variables. The rows correspond to the cases or observations. The data are coded as follows;

ID: A unique number to identify a case

AGE: Age in years

EDUC: 1: Primary school

2: Some secondary school

3: leaving certificate

4: Some College

5: College Degree

6: Post graduate degree

Gender: 1: Males

2: Females

EARN: Total annual income

Job class: 5: Private Sector

6: Public sector

7: Self Employed

Type and run summary(Lab1$AGE) to see a selection of summary statistics for the variable AGE.

Type and run table(Lab1$Gender) to see the frequencies of each gender category.

Type and run table.gender <- table(Lab1$Gender) to create a table in memory named table.gender.

Type and run prop.table(table.gender) to see the proportions of each gender category.

Type and run table(Lab1$EDUC, Lab1$Gender) to see a two-way cross-tabulation of the frequencies of variables EDUC and Gender.

Type and run ftable(Lab1$EDUC, Lab1$Gender, Lab1$Job.class) to see a three-way cross-tabulation of the frequencies of variables Educ, Gender and Job.class.

Type and run hist(Lab1$AGE) to see a basic histogram of the variable AGE.

Type and run boxplot(Lab1$AGE) to see a basic boxplot of the variable AGE.

Type and run boxplot(Lab1$AGE~Lab1$Gender) to see a basic boxplot of the variable AGE in each level of the variable Gender.

Type and run plot(Lab1$EARN, Lab1$AGE) to see a basic scatterplot of EARN and AGE.

Type and run Lab1$EARN\_1000 = Lab1$EARN/1000 to create a new variable equal to EARN divided by 1000.

Type and run plot(Lab1$EARN\_1000, Lab1$AGE) to see a basic scatterplot of Lab1$EARN\_1000 and AGE.