

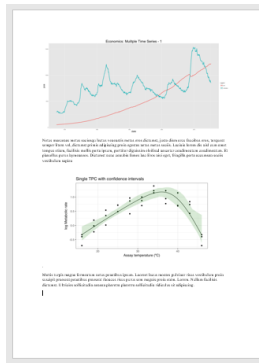
**Assignment 2: Data visualisation using R**  
**Due: 18<sup>th</sup> September 2020 Friday @ 6pm**

For **each task** you are required to submit:

- Question X.pdf



or



```

1 # Import the modules
2 import cv2
3 import numpy as np
4
5 # Load the image
6 img = cv2.imread('img.jpg')
7
8 # Convert the image to grayscale
9 img_gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
10
11 # Apply thresholding
12 ret, thresh = cv2.threshold(img_gray, 127, 255, cv2.THRESH_BINARY)
13
14 # Find contours
15 contours, _ = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
16
17 # Draw contours
18 cv2.drawContours(img, contours, -1, (0, 255, 0), 2)
19
20 # Display the image
21 cv2.imshow('Contours', img)
22
23 cv2.waitKey(0)
24 cv2.destroyAllWindows()

```

- 1

- This assignment forms **40% of your total assessment** for the subject. (The total mark that you achieved out of 100 will be converted into a mark out of 40 for your overall unit grade.)
- The data sets needed for this assignment are **located on the LMS in CSV format**. The LMS also contains a **data-dictionary** for some of the given data sets. (Refer to this so you know what the variables in the data represent).

#### **OTHER IMPORTANT INFORMATION**

- Standard plagiarism and collusion policy, and extension and special consideration policy of this university apply to this assignment.
- A cover sheet is **NOT** required. By submitting your work online, the declaration on the university's assignment cover sheet is implied and agreed to by you.

### TASK 1 [25 marks]

Using the data contained in the file `covid_au_state.csv`, create a visualisation that:

- Shows the total number of new confirmed cases daily throughout Australia since the first cases recorded on 25 January 2020. Highlight the three days with the highest number of new cases.
- Compares the daily growth factor across all the eight states in Australia from 17 March 2020 to 16 August 2020. (The growth factor is calculated by dividing the new cases on the current day with the new cases of the previous day.)

[Data source: [covid19data.com.au.](https://covid19data.com.au/)]

### TASK 2 [25 marks]

Using the data contained in the file `grand_slam_data.csv`, create a visualisation that:

- Compares the performance all the winners for each of the four Grand Slam tournaments (Australian Open, French Open, Wimbledon, U.S. Open); and highlights the top winner for each tournament.
- Shows and highlights how many times the winners and the runner-ups met each other on each of the Grand Slam tournament finals for the 40 most recent tournaments (from 2008 to 2017).

### TASK 3 [25 marks]

Using the data contained in the file `big_stock_data.csv`, create a Shiny interactive data visualisation that allows the user to:

- View the change over time of a chosen share based on either the closing price or the share volume.
- Compare the shares performance between two or more companies.

[Note: You only need to describe one screenshot for this task – as we will run your R code to check the interactivity of your visualisation.]

### TASK 4 [25 marks]

Using the data contained in the file `crime_data.csv` and the **ggmap** R package, create a Shiny interactive data visualisation that allows the user to:

- View the spatial density of one or more offense types in the Houston area.
- View the spatial density of all the crimes happened for a time period (e.g. from 01/03/2010 to 30/04/2010).

[Note: You may want exclude the offense type of Theft when creating the visualisation. You only need to describe one screenshot for this task – as we will run your R code to check for the interactivity of your visualisation.]

## ASSESSMENT CRITERIA

Assessment criteria	A (80-100%)	B (70-79%)	C (60-69%)	Pass (50-59%)
Ability to describe elements of data visualisation	All visualisations are described clearly and concisely. Descriptions include information about all relevant visual elements.	Some visualisations are described clearly and concisely. Descriptions include some information about the relevant visual elements.	Limited ability to describe visualisations. Limited description of visual elements.	The visualisations have a description. Descriptions briefly refer to visual elements.
Ability to create data visualisations in R	All visualisations are created appropriately and without error. All data files are imported in correct format.	Some visualisations created without error. Some data files imported with no errors.	Limited ability to create visualisations. Some data imported with errors.	Able to identify data visualisation tools. Able to identify data importing tools.
Correct visualisation theory applied to charts and ability to convey information through data visualisation	All visualisations are informative, efficient and beautiful. All visualisations clearly convey the appropriate message.	Most visualisations are informative, efficient and beautiful. Some visualisations convey the appropriate message.	Some charts are informative, efficient and beautiful. Limited ability to convey messages through data visualisation.	Able to identify appropriate chart type. Able to identify appropriate message.