Lab Assignment 1

Exploring the world of Big Data

The internet of things opens up a host of opportunities, but also brings about significant challenges - one of which is privacy. As we move through the modern world many of our actions leave behind an electronic footprint; who collects that data and what they do with it have become critical questions for modern society.

Australia's new data retention laws mean phone and internet companies have to save this information for two years: that's every time you call someone, where you call them from, which cell tower your phone pings every time it connects to the internet, and more.

What's the size of your electronic footprint? How much data is collected and stored as you go through your day? On a mission to find out what that data might reveal, ABC reporter Will Ockenden took a 'surveillance selfie': he got access to his own metadata, and in this assignment you will see what an individual Australian's metadata actually looks like, and what you can infer from it.

In the debate over mandatory data retention, the Federal Government and security agencies repeatedly said there was nothing to worry about.

"We're talking here about metadata; we're not talking here about the content of communications," Prime Minister Tony Abbott said. "It's just the data that the system generates."

But metadata still says a lot about your day-to-day life.

The dataset Will Ockenden received contained a year's worth of outgoing call and SMS records, and six months of his 'data sessions', which are the records kept every time his phone connected to the internet over the mobile network.

All in all, this simple data request returned 13,000 individual records. There were 1,500 outgoing phone calls and SMSes but the vast majority - 11,200 records – were data sessions, complete with the time and date his phone connected to the mobile network and which cell tower it connected to.

In other words, by carrying a smartphone Will was in effect carrying a tracking device that logged roughly where he was every 20 minutes of every day, on average.

Government departments, police and security agencies have access to all the data Will received about himself - and more - without the need for a warrant.

Being able to follow someone's daily movements is one thing but it's once we start to collate and visualise the data that patterns can start to emerge.

What's in the data?

The dataset Will Ockenden received from Telstra included:

- Who he called and texted (in our dataset, exact phone numbers have been hidden and replaced by unique identifying codes).
- How long each phone call lasted.
- The time of the communication.
- The location of the cell tower contacted when outgoing calls were initiated.
- The location of the cell tower contacted for SMS and internet connections.

Other data that is to be kept by telcos and internet providers under data retention laws – but which is not included in the dataset released to Will – includes but is not limited to:

- Details of incoming phone calls.
- The time, date, size and recipients of emails sent using your ISP's email service (i.e. not via webmail services such as Gmail, Yahoo and Hotmail).
- The file type and size of any attachments sent or received with emails, when using your ISP's email service.
- Details about internet usage including how much bandwidth the internet service provides.

Assignment

- Browse to : https://public.tableau.com/en-us/s/download and download Tableau public after creating a free account.

- Then download the data provided in wills_metadata.csv

- Watch the Lab Assignment 1 video and answer the four questions

- Upload your answers to the four questions along with a screenshot of each of the visualisations you created.