Making Sense of Crim Data

Reka Solymosi

2018-09-23

Contents

Introduction

This workbook contains the lab materials and homework assignments for an introduction to data analysis course designed for LAWS20441 Making Sense of Criminological Data, a 2nd year undergraduates on the BA Criminology programme at the University of Manchester.

It makes use of Excel, as we have identified a gap in training students to use Excel, despite it being a primary tool for data analysis (whether we like it or not) in many public and private sector organisations. As many students take Q-step internships, this skill was identified as important.

Making Sense of Crim Data introduces students to data, and the concepts of descriptive data analysis. The role of this term is to familiarise students with basic concepts of data analysis, and get aquainted with descriptive statistics to be able to talk about data about crime, policing, and criminal justice topics. Details can be found in the Syllabus.

0.1 Disclaimer

Please beware that:

- In making these notes, while I briefly cover some concepts, students are expected to do the weekly reading, and attend the weekly lectures, as well as participate in lab disucssions to receive a complete course experience. These notes are *not* intended to be a stand-alone reference or textbook, rather a set of exercises to gain hands-on practice with the concepts introduced during the course.
- These pages are the content of the BA Criminology 2nd year course Making Sense of Criminological Data. They are meant to (very gently) introduce undergraduates to the concept of data analysis, and cover descriptive statsitics and the key concepts required to build an understanding of quantitative data analysis in crime research. It is followed in the second term by Modeling Criminological Data where students cover inferential statistics. The notes presented here are supported by compulsory reading and some lectures, and so do not provide a comprehensive description of these techniques and tools and how to use them.
- The handouts below use, among other data sets, dara from the UK data service such as the Crime Survey for England and Wales that is available under a Open Government Licence. This dataset is designed to be a learning resource and should not be used for research purposes or the production of summary statistics.

0.2 Overview of course

0.2.1 Module structure

Hi there and welcome to this course for making sense of criminological data. This introduction will explain the structure of the course.

The course is 10 weeks, each week made up of 5 elements:

1) Preparatory reading
2) Lab session
3) Post-lab task
4) Homework quiz
5) Lecture

0.2.1.1 1) Preparatory reading

For each week you will receive some preliminary reading or videos to watch, before coming to the session. It is very important that you read these before coming to the lab session, as it will make engaging with the lab material easier. Also, you can take the labs as an opportunity to ask questions about the readings, and discuss with myself and the teaching assistants during the 2-hour lab sessions.

0.2.1.2 2) Lab session

Lab sessions are the two-hour sessions in computer labs, where you are to work through the lab notes included in this book. You should take time to engage with these notes, and ask lots of questions from myself and the teaching assistants present. This is a time to really engage with the materials.

When you come into a lesson, you will be able to sit at a PC and get started straight away. You will find the instructions for each week in this booklet. You can open up the link via Blackboard, and read through the instructions chronologically. That just means start at the top, and read through to the bottom.

Usually within the notes there are some general introduction to the topic covered that week, with links to videos or further reading. You should come equipped with headphones to watch the videos.

These lab notes also contain within them activities. Activities will be denoted by red colour text, just like this one You should do these activities in the lab, and ask for our help when you are stuck, or if you do not understand a concept. These activities will help you with your learning, but also will contribute towards your homework. You are welcome to discuss with each other, and with us, but please do make sure that when it comes to understanding the learning behind these activities, you are confident in your ability. The final essay will rely heavily on your ability to take the concepts you learn duing the activities, and apply them in a way that shows your understanding.

0.2.1.3 3) Post-lab task

Each week, after you have completed the lab notes, you must complete some post-lab tasks. These will take the form of a worksheet. You can find each worksheet and relevant material (eg: data) on Blackboard in the folder for that week. There will also always be a link at the bottom of the lab notes. You have to complete these tasks in order to be able to take the homework quiz (which is assessed). The tasks will always mirror the in-lab activities, so if you get through those, the task should be a breeze.

0.2.1.4 4) Homework quiz

Each week you will have to complete a homework quiz. This quiz is assessed, and your score on all the quizzes combined counts for 20% of your final mark (the other 80% is your final essay). The questions in the homework quiz will ask about key concepts from your reading, and about the answers to the post-lab task. Make sure that you have finished the task before you begin your homework quiz, and have it with you while you do so. The homework quiz will be available on Blackboard, and will be open on each Friday at 11:00am, and close the following Thursday at 9:00am. You can take the homework quiz any time between these times. You can take it only once. Once you activate the quiz, you will have only 30 minutes to complete it. Please make sure you are in a quiet environment where you will not be disturbed, with your reading notes and

your post-lab task with you, so you can complete the homework quiz successfully. Upon submission you will receive immediate feedback.

Don't forget your homework is graded, and counts towards your final mark. But you get to practice for completing the tasks with the activities in the lab. And if you have time left over, you can always complete these tasks here in the lab. No matter where you do them, by havign the tasks completed and with you when you take the quiz, you will be more confident in the homework quiz exercises.

0.2.1.5 5) Lectures

In the lectures we will go over one last time the basic concepts covered in that week. The lecture is the final event that should pull together all your learning from that week. It is your chance to ask questions, discuss, and further interrogate the material we cover. I encourage you to bring your own examples to lectures wherever you encounter them. Lectures are podcast, however they are also attendance monitored. You must attend, and sign in to these lectures to receive attendance credit.

Chapter 1

Week 1

1.1 Learning outcomes

Welcome to week 1! This week is all about getting set up, and taking some first steps in learning about what is *data*. These are the building blocks of all the weeks to come, so pay special attention, and do set a pattern of asking questions! Now without further ado, let's get started.

1.2 Setting up your working environment

There is a myth about the scientist and the messy workspace, typically illustrated with Albert Einstein:
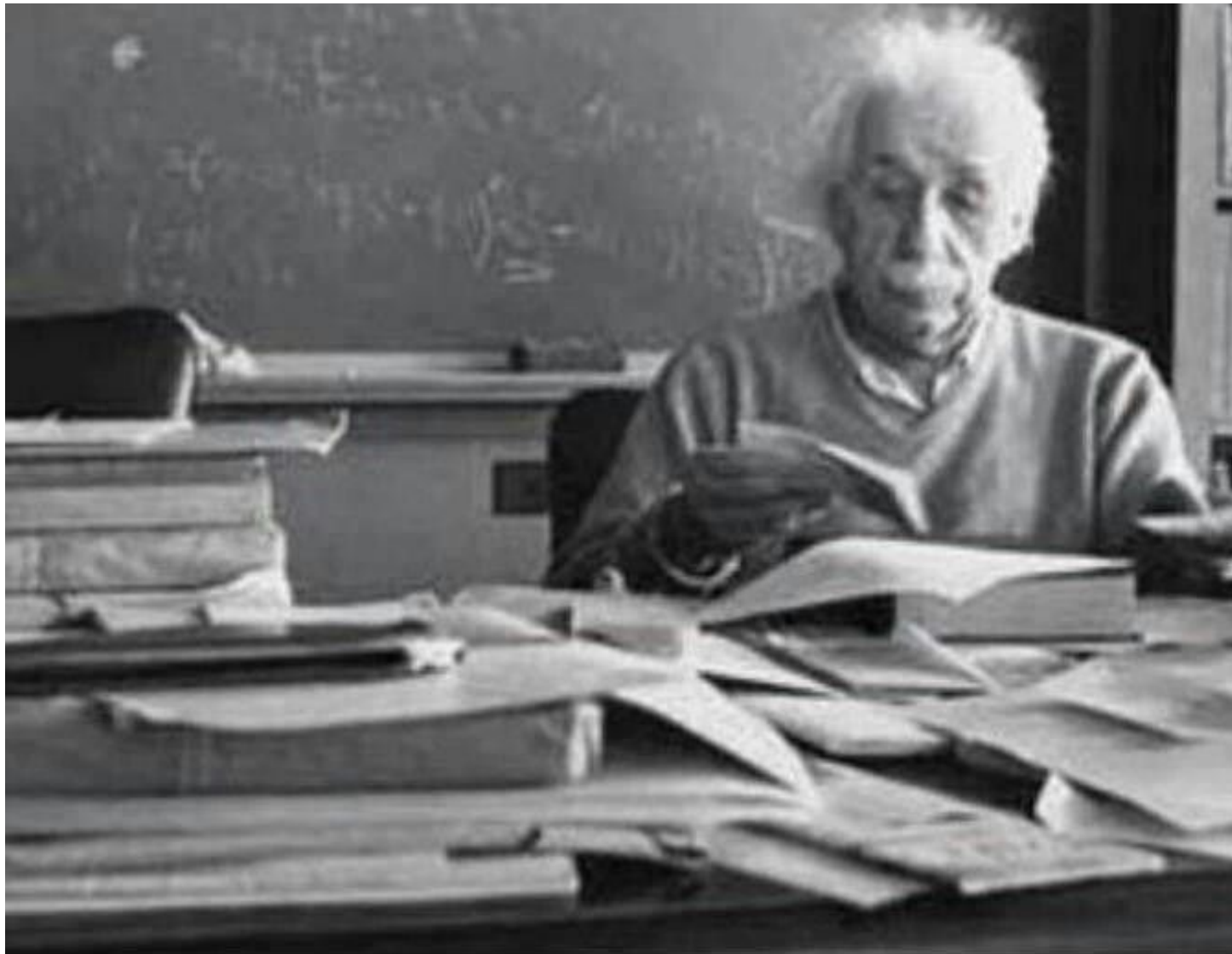
"If a cluttered desk is a s
cluttered mind, of what,
an empty desk a sig

— Albert Einstein

However many of us need order to be able to work properly. An organised workspace is also prominent, as we can see with these famous work spaces of : Galileo, Marie Curie, John Dalton, Alan Turing, and Charles Dickens:

When working with data, you have to consider your workspace. You can think of your computer folders as your desk. It helps immensely to keep our data, your code and your notes organised. You will likely have a project folder, where you save your data, your graphs, your analysis outputs, etc. You want to consider the layout of this folder, how many subfolders will you have, what is the best structure to work for you? You might think this is trivial, but when you are working on a project with multiple data sets, or many graphs, it can get very messy very quickly. I recommend going through this resource from the university of Cambridge data management guide to consider Naming and Organising Files, Documentation and Metadata, Managing References, and Organising E-mail.

Normally, say when you are working on a Windows PC at home, you can save files to your C: drive, and easily access them the next time you turn the computer on. When you are using computers at the university, however, you cannot be certain of being able to use the same computer again when you want to find and use your file again. Fortunately, once you have logged on to a computer at the university, a personal drive is available for you to use, and in effect, this drive follows you around to whichever PC you are logged on to. All students and staff have a personal file storage space on the network - known as the P: drive as this is usually the network drive letter allocated to it. Wherever you log on to the campus network your P: drive is available (in PC Clusters the My Documents icon on the desktop is a shortcut to the P: drive).

You should already have a username and password with which to log on. Your username has 7 or 8 letters/digits, typically beginning 'm...'. On the computers in the Faculty clusters (where you work in class time) you will see the Faculty computer 'image' and a version of the operating systems Windows 7.

Away from the campus you can download and upload files to and from your P: drive over an Internet connection - for example to and from your home computer, but please remember that large files may take a long time to upload/download depending on the speed of your home internet connection. See here for details about remote access.

It's generally good to create a folder to save your data and our outputs in, which we will call a **working directory**. So firstly, before we begin to do any work, we should create our **working directory**. This is simply a folder where you will save all our data, and also where you will be reading data in from. You can create a new folder, where you will save everything for this course, or you can choose an existing folder. It's advised that you create a folder, and also give it some name you remember, that will be meaningful. Generally try to avoid spaces and special characters in folder (and file) names. Here is a handy guide you should read about naming files and folders that will be relevant for all your future work.

It's not necessarily a good idea to just dump everything into 'Desktop' either, as you want to be able to find things later, and maybe keep things tidy. We will now show you where your personal drive is and how to save your word file to it.

Just to have some practice with doing this, create a Word document to put into the folder. Open Microsoft Word now (you can find this on the 'Start' button in the bottom left corner of your screen). The only thing you have to type in the document is your name.

Save the file to your p: drive as follows. Click on the *save* button (or click File then Save in the drop down menus). The dialogue box below will appear. Click on the down arrow in the 'Save in:' box to change the default directory to your P: drive. The P: drive is the one that starts with your user name. Click on this drive.
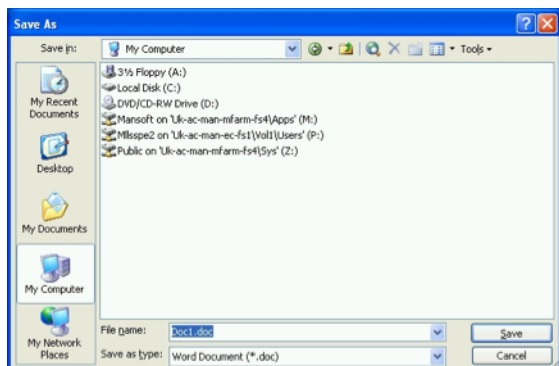
Now create a new folder within your P: drive, especially for material for this course. To do so, click on the 'Create new folder' icon, as shown in the dialogue box below.
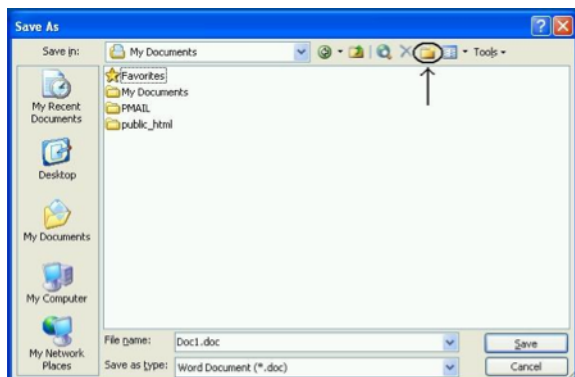
Label the new folder 'LAWS20441'. You will now have a series of folders in your p: drive (some of these you haven't created yourself, they have been provided for you by the university), one of which you can store course material in. Finally, name your Word document 'Trial Document' (in the filename box) and click 'Save'. As noted above, you can remotely access your p: drive (from home or elsewhere). You can do this by logging into your personalised university portal (https://my.manchester.ac.uk).

A word of caution, if your P: drive is full (and this tends to happen when you save image or sound files to it), there is a chance that some of the applications you want to use do not work. So make sure you keep your P: drive tidy if you don't want to run into problems.

1.3 Getting to know Excel

Besides having a folder where you save something, you should also get comfortable with the tools which you will be using throughout this course. The main tool you'll be using is Microsoft Excel. You will be using it to explore, learn about, and manipulate criminological data throughout this course.

If you are following along on your own laptop (I encourage this strongly) you can download microsoft excel **for free**, courtesy of the UoM library. Follow the instructions here to get Microsoft Office on your laptops.

Now you likely have come across Excel before, but it's also possible that you have not, so I will start with the assumption that this is your first time opening it up. Exciting. So let's get to it. To open excel, click on the windows icon, and select Microsoft Excel from the Microsoft Office bundle.
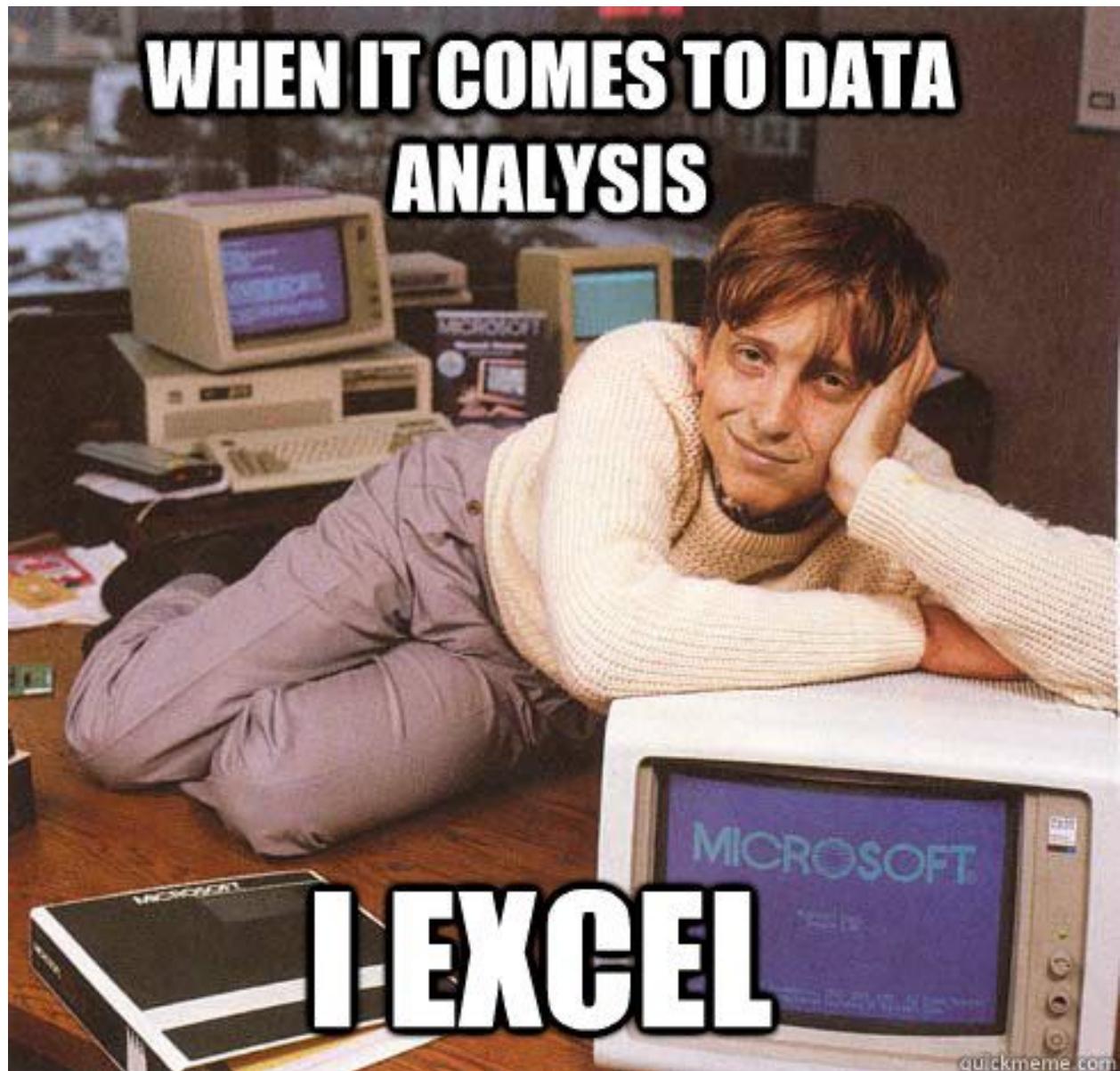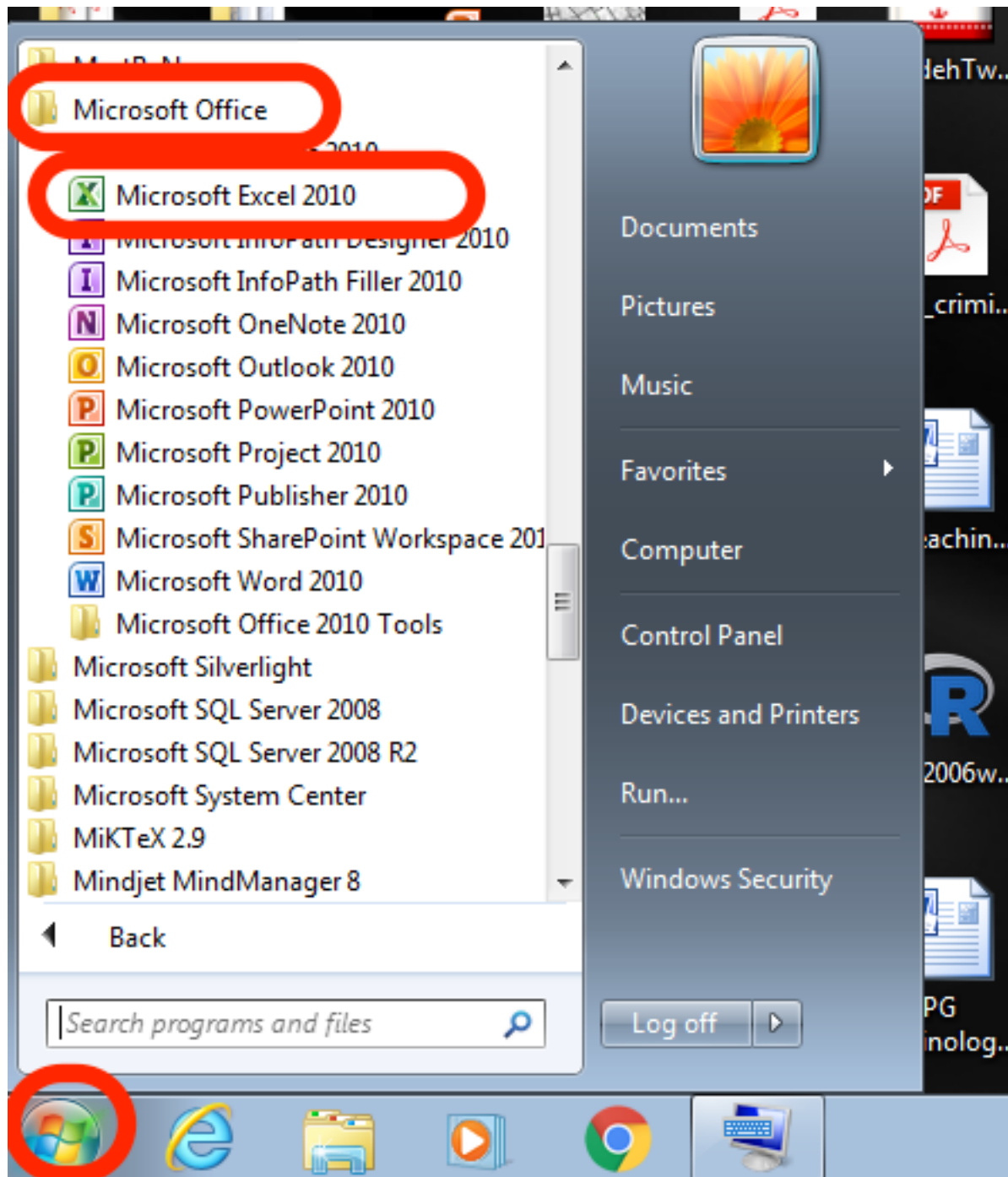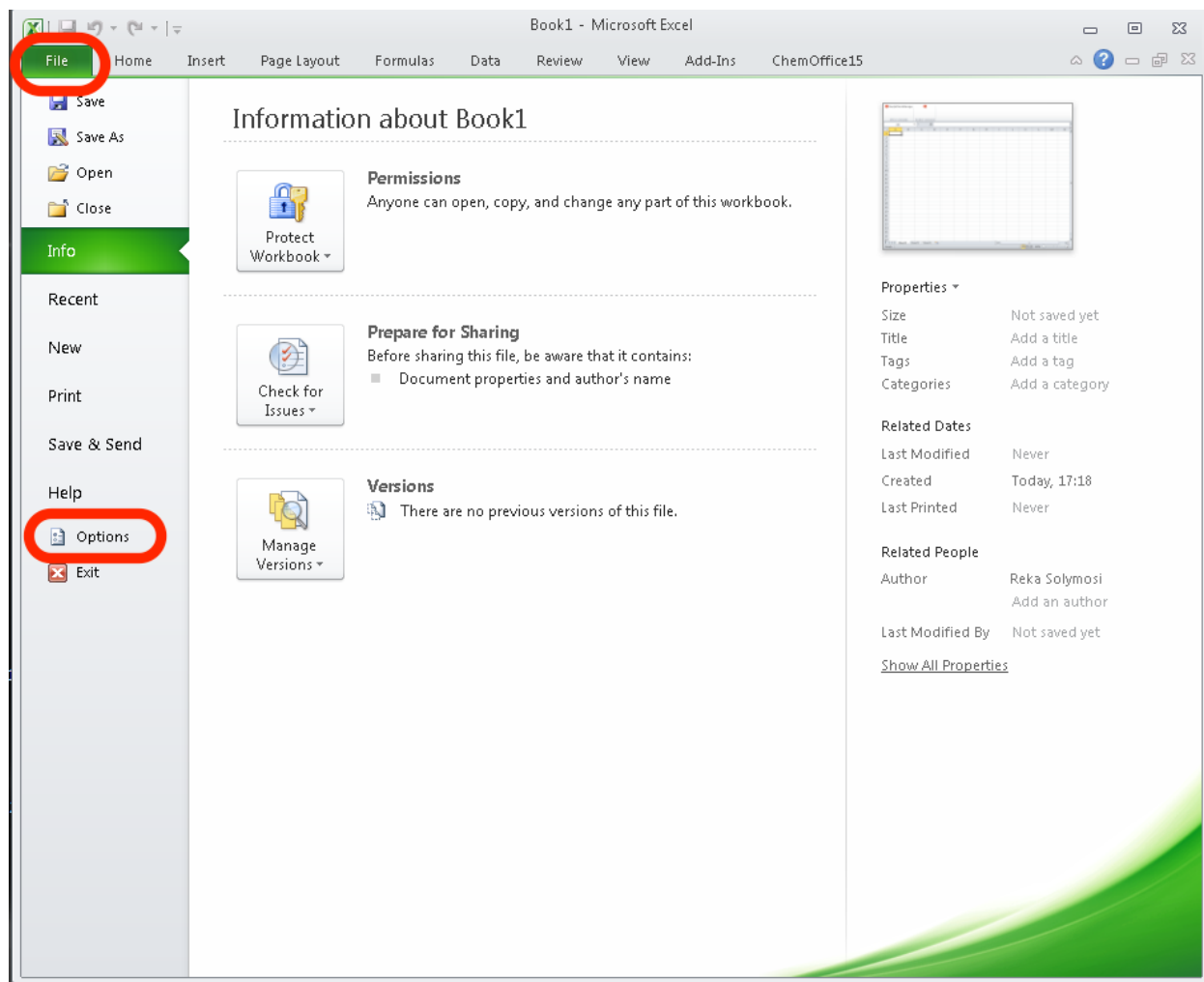
Figure 1.1: I excel

Now you should have Excel open, and you will see an empty spreadsheet. We will be using this just a little bit later on. But for now there is one more step we need to do, to be fully set up. We need to install the data analysis toolpak.
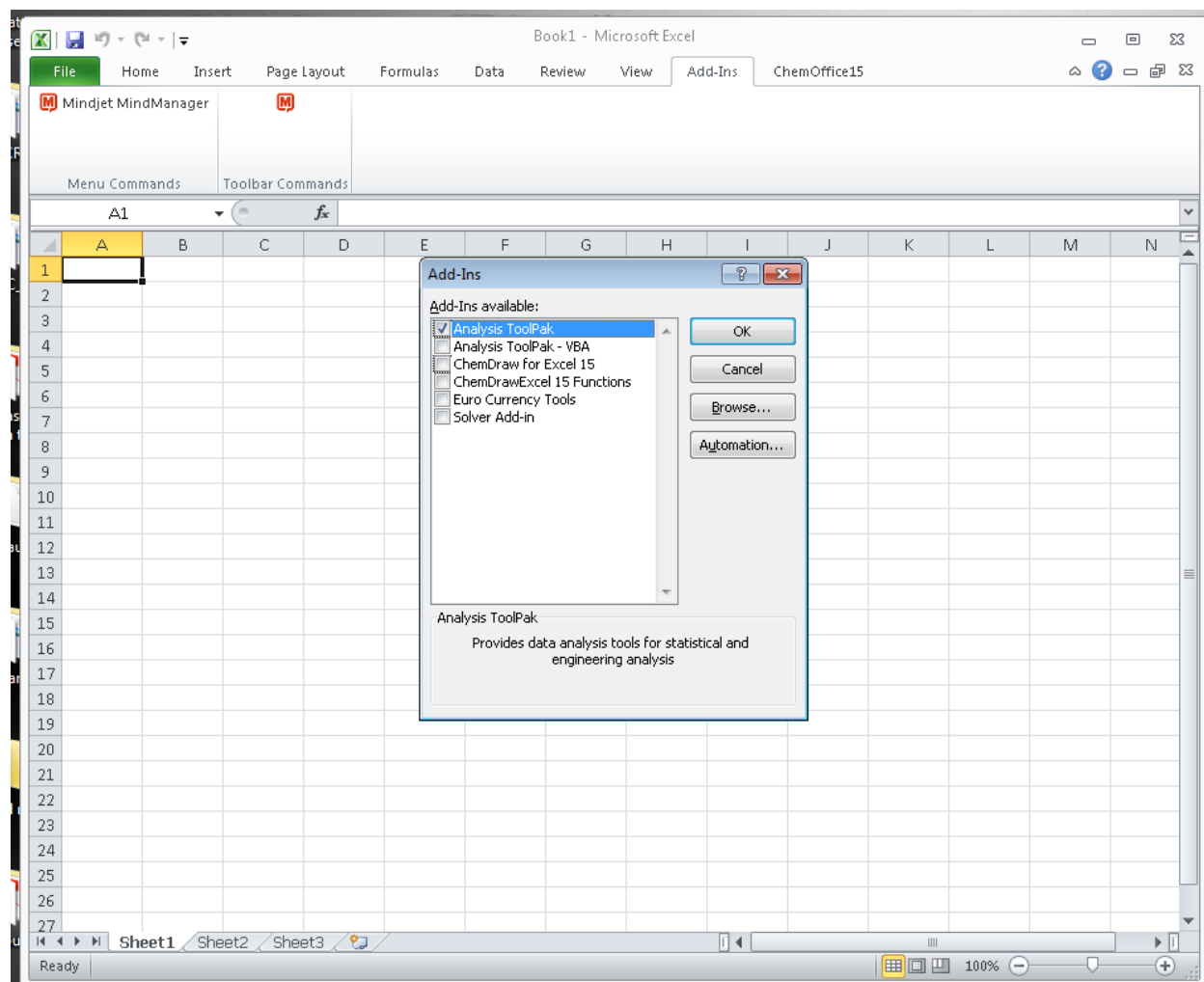
To do this, click on the 'File' tab, and click on 'Options':

This will bring up a popup window. Here click on 'Add-Ins', then highlight 'Analysis Toolpak' and click on 'Go':
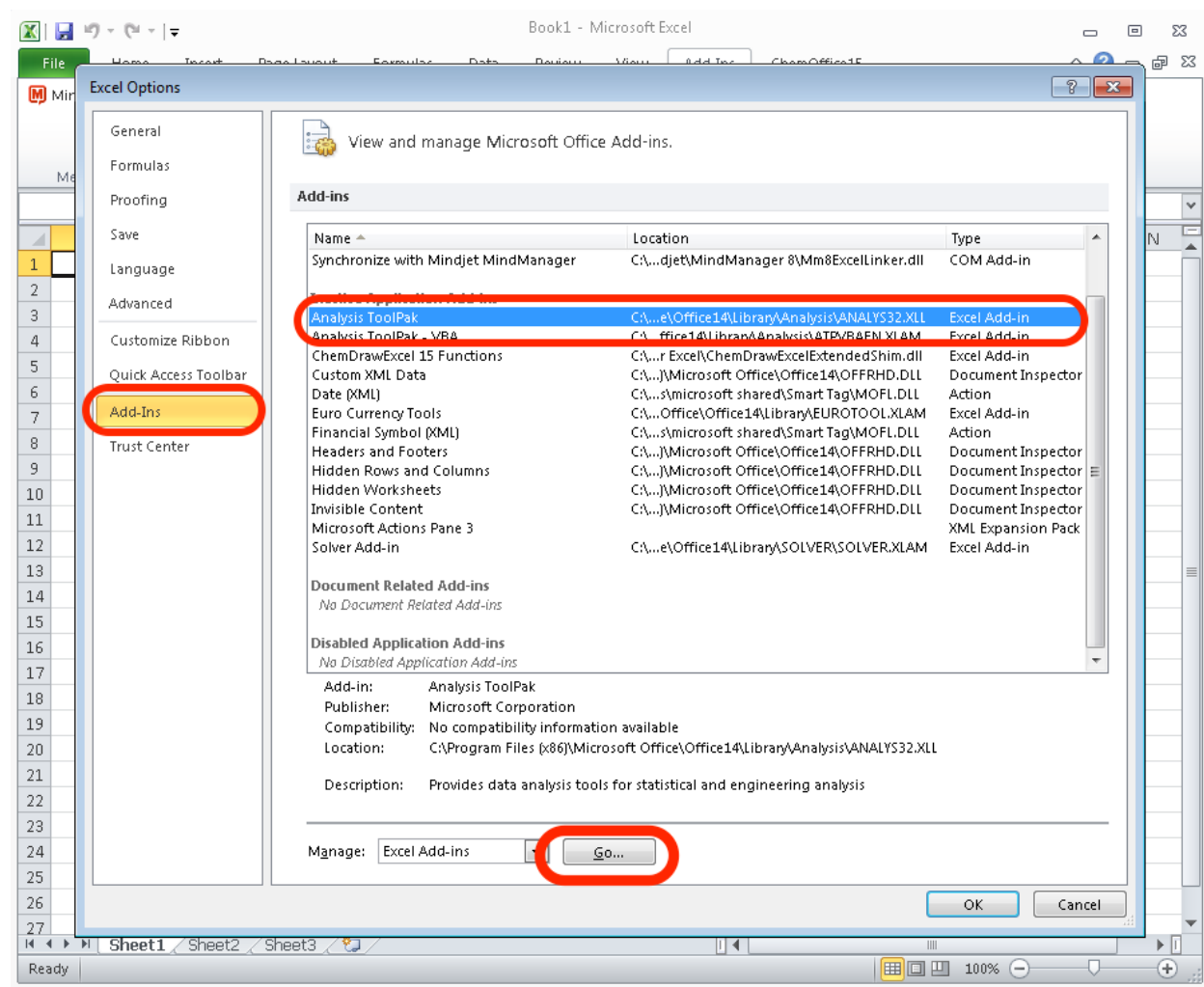
This will open another popup window. Here make sure you tick the box next to 'Analysis Toolpak' and click 'OK':

Click 'OK' and you should be done! You can check by clicking on the 'Data' tab, and checking to see if a little Data Analysis icon has appeared:
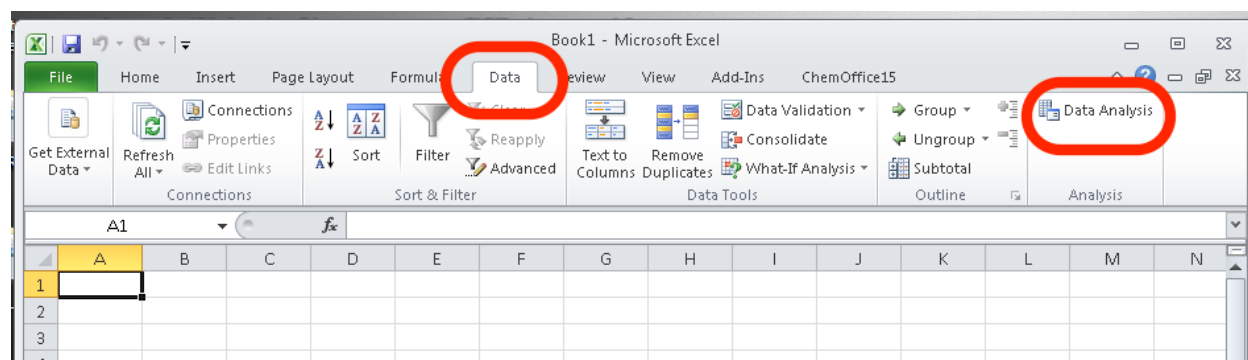
If you are confused, see here for instructions how to get this. Once you have successfully installed the data analysis toolpak it will appear.

And that's it you are now set up! Excellent!

You can now move on to the subtantive part of today's course. In the next section we will learn about variables and data.

1.4 Data: Variables and observations

We know that in the period from May 2016 to May 2017, Greater Manchester police recorded a total of 420228 crimes. We also know that the largest number were recorded in Anti-social behaviour crime categeory, with 122443 instances, while the fewest in Possession of weapons, with 2009 instances.

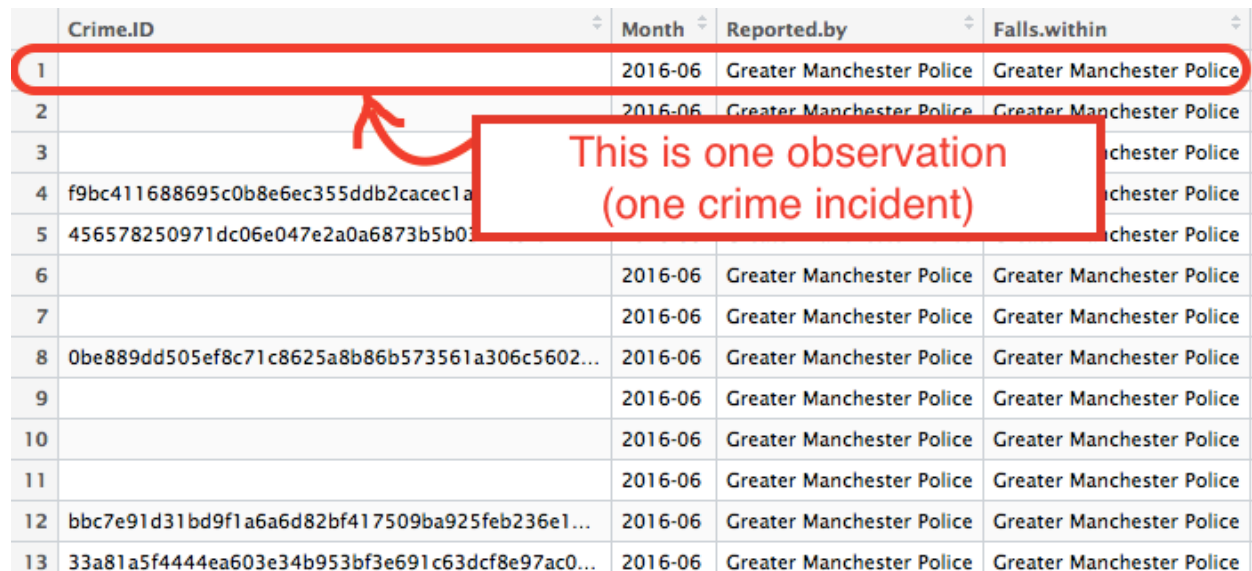We can also track changes in the number of crimes over time:
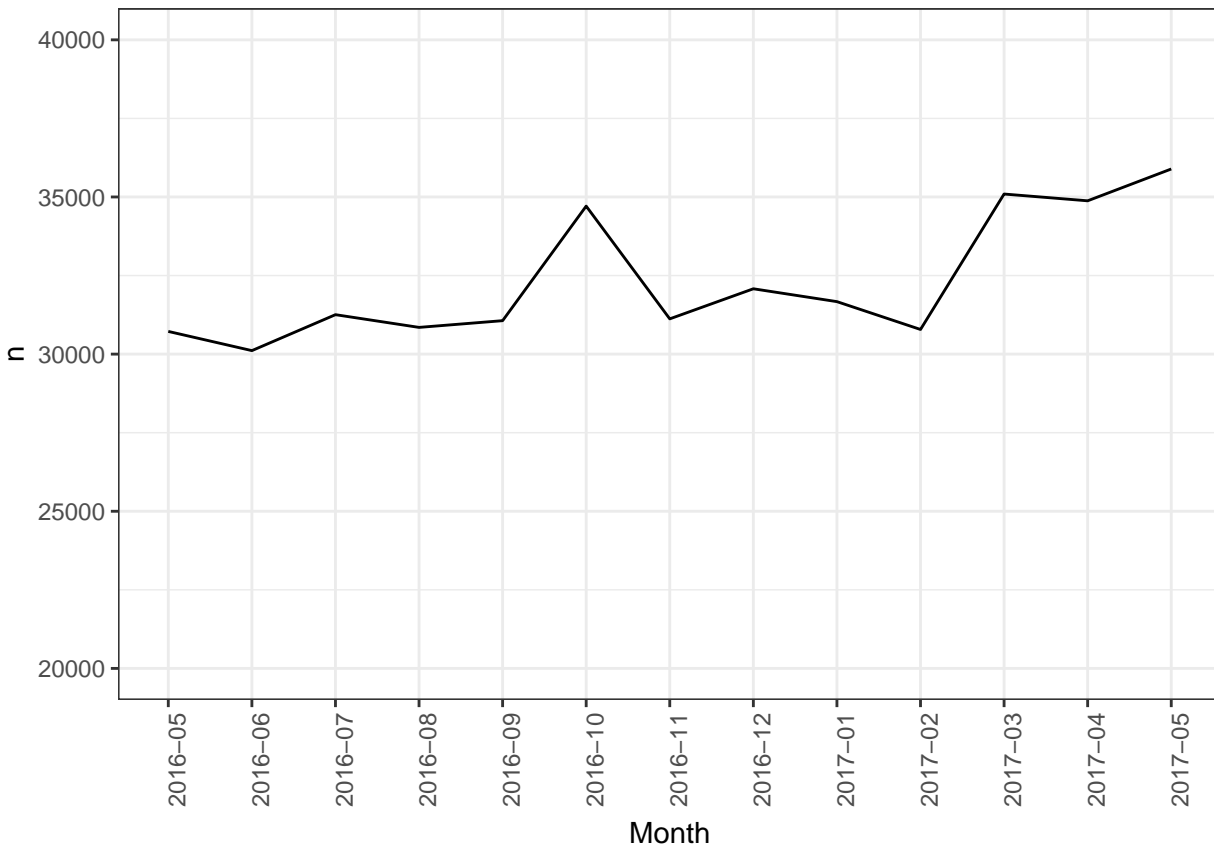
How do we do this? Well, in the United Kingdom, since 2011 data regarding individual police recorded crimes have been made openly available to the public via the police.uk website 1. This means that by visiting the data.police.uk/ website you can access data about street-level crime, outcome, and stop and search information, broken down by police force. What does this mean? What do these data look like? Let's have a look:

X	Month	Reported.by	Falls.within	Longitude	Latitude	Location
1	2016-05	Greater Manchester Police	Greater Manchester Police	-2.462774	53.6221	On or near Scout Road
2	2016-05	Greater Manchester Police	Greater Manchester Police	-2.464422	53.6125	On or near Parking Area
3	2016-05	Greater Manchester Police	Greater Manchester Police	-2.464422	53.6125	On or near Parking Area

In this data set each row is one crime record:

For every single crime event recorded in this data, there is a row, and it contains all the information that we know about this crime incident. It will have a value for each variable that we are interested in. The variables are the columns.

So for example, *month* is a variable in our data, and for every row (which is every crime incident) this variable can take a value. Every crime incident occurred at one specific month, and that month when each incident happened will be the value that the *month* variable will take. And the month column will contain

	Crime.ID	Month	Reported.by	Falls.within
1		2016-06	Greater Manchester Police	Greater Manchester Police
2		2016-06	Greater Manchester Police	Greater Manchester Police
3				chester Police
4	f9bc411688695c0b8e6ec355ddb2cacec1a			nchester Police
5	456578250971dc06e047e2a0a6873b5b0...			nchester Police
6		2016-06	Greater Manchester Police	Greater Manchester Police
7		2016-06	Greater Manchester Police	Greater Manchester Police
8	0be889dd505ef8c71c8625a8b86b573561a306c5602...	2016-06	Greater Manchester Police	Greater Manchester Police
9		2016-06	Greater Manchester Police	Greater Manchester Police
10		2016-06	Greater Manchester Police	Greater Manchester Police
11		2016-06	Greater Manchester Police	Greater Manchester Police
12	bbc7e91d31bd9f1a6a6d82bf417509ba925feb236e1...	2016-06	Greater Manchester Police	Greater Manchester Police
13	33a81a5f4444ea603e34b953bf3e691c63dcf8e97ac0...	2016-06	Greater Manchester Police	Greater Manchester Police

This is one observation (one crime incident)

Figure 1.2: Each row is one observation

all the instances of the month variable for each crime incident recorded. Each observation (crime) will have a value for this variable (the month that it was recorded).

Let's have a go at recording some data observations and putting them into a database, to give you some hands on experience here.

1.4.1   Activity 1: Building your own data

You will get a better understanding of how data represents what you are measuring if you have a go at building your own data set. We will do this here by using data from twitter. You are most likely familiar with twitter. You probably even tweet yourself. But even if you have never used twitter, you will no doubt know someone who does. In fact, many police forces use twitter. GMP is one of these forces, and in particular, GMP city centre like to keep their followers updated. Recently, the MEN had an article based on following GMP city centre's tweets for one Saturday night. You can read about that here.
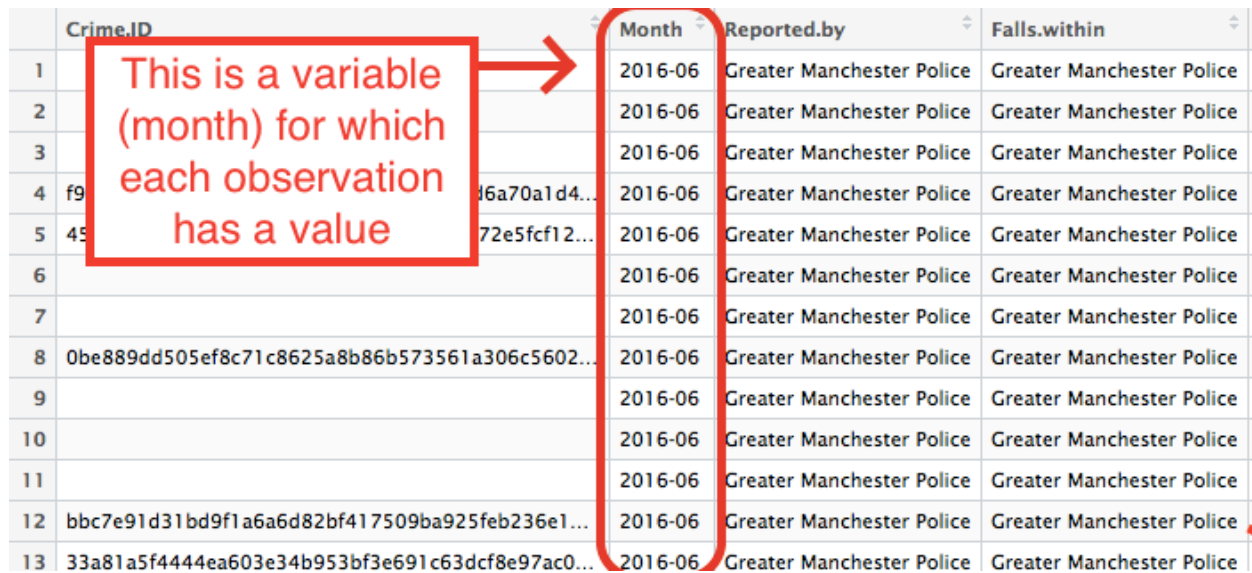
Evidently tweets present really exciting and rich data. However they do not come in a format that is readily available for analysis in the form that we just presented here. But what you can do is collect data from tweets. And this is your task for your first lab activity.

I have collected for you a set of tweets. Your task is to turn this into a rectangular data format, with the columns as variables, and the rows as observations (tweets). Let's go through how to do this, setp-by-step.

So first things first, we need a tool. As discussed we'll mostly be using Excel in this course. So open up excel and create a brand new spreadsheet.

Your first activity is to create a column header for each variable we want to collect. The easiest way to do this is just to make the first row your column headers. You can go ahead and create a column for each of the variables we are interested in collecting about each tweet. These are:
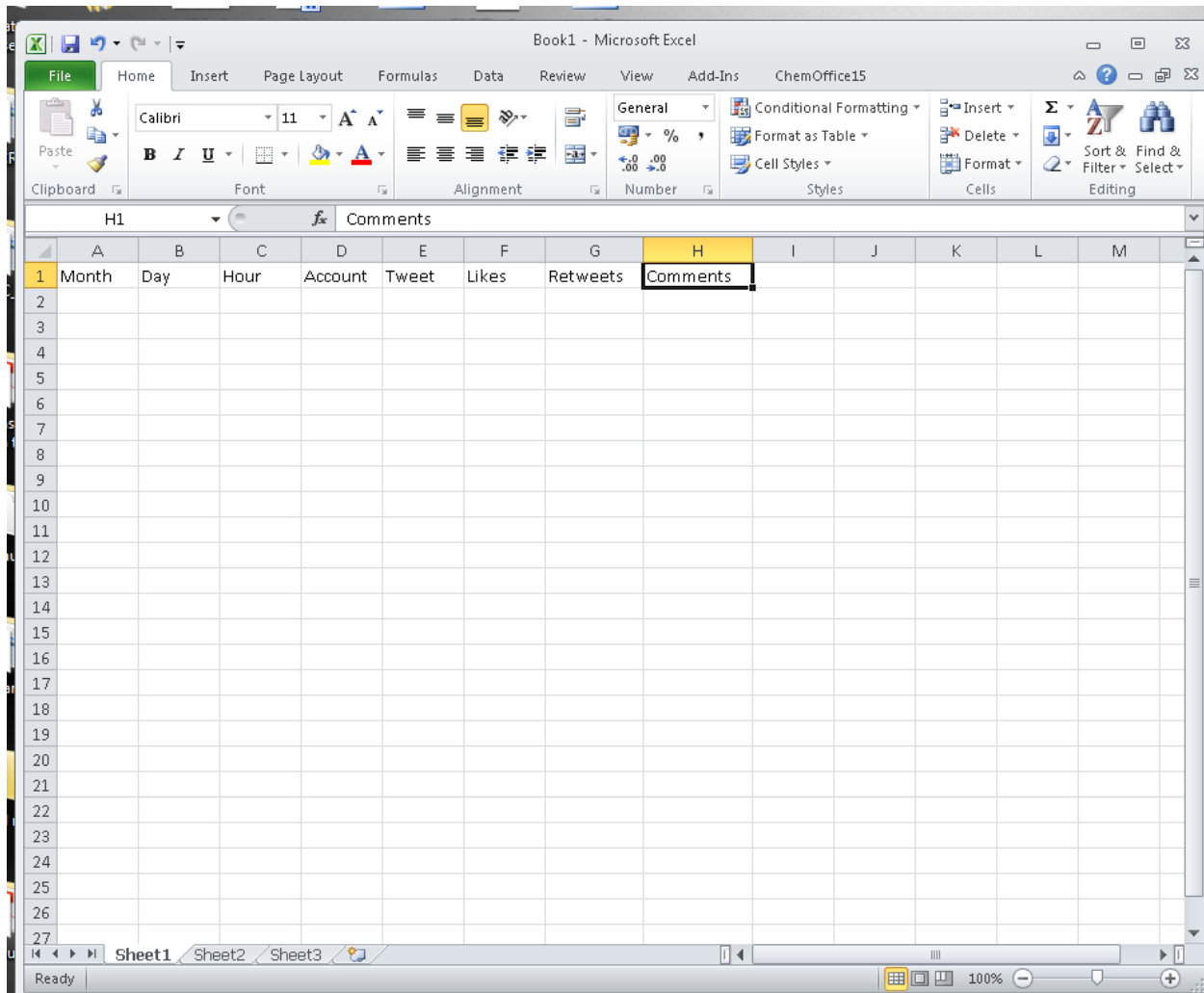
- Month: The month in which the tweet was sent
- Day: The day of the month in which the tweet was sent
- Hour: The hour when the tweet was sent, in 24h format (where 13:00 is 1pm and 01:00 is 1am)
- Account: The account who tweeted this tweet
- Tweet: The content of the tweet

Figure 1.3: Each column is a variable

- Likes: Number of likes for this tweet
- Retweets: Number of times this tweet was retweeted
- Comments: Number of comments made as reply to this tweet

Now you will just have to create a new row for each tweet, and populate a value for each variable we are collecting in our data. I'll go through the first tweet with you, so we're clear on what's happening. This is tweet number 1:

- Tweet 1

You will see this open in a new window. Now let's try to find the value for each variable in this tweet:

- *Month*: July
- *Day*: 31
- *Hour*: 06
- *Account*: GMPCityCentre
- *Tweet*: Man left £1000 Stella McCartney bag on seat in Village bar with person he had just met, and when returned, woman and his bag had gone
- *Likes*: 14
- *Retweets*: 43
- *Comments*: 20

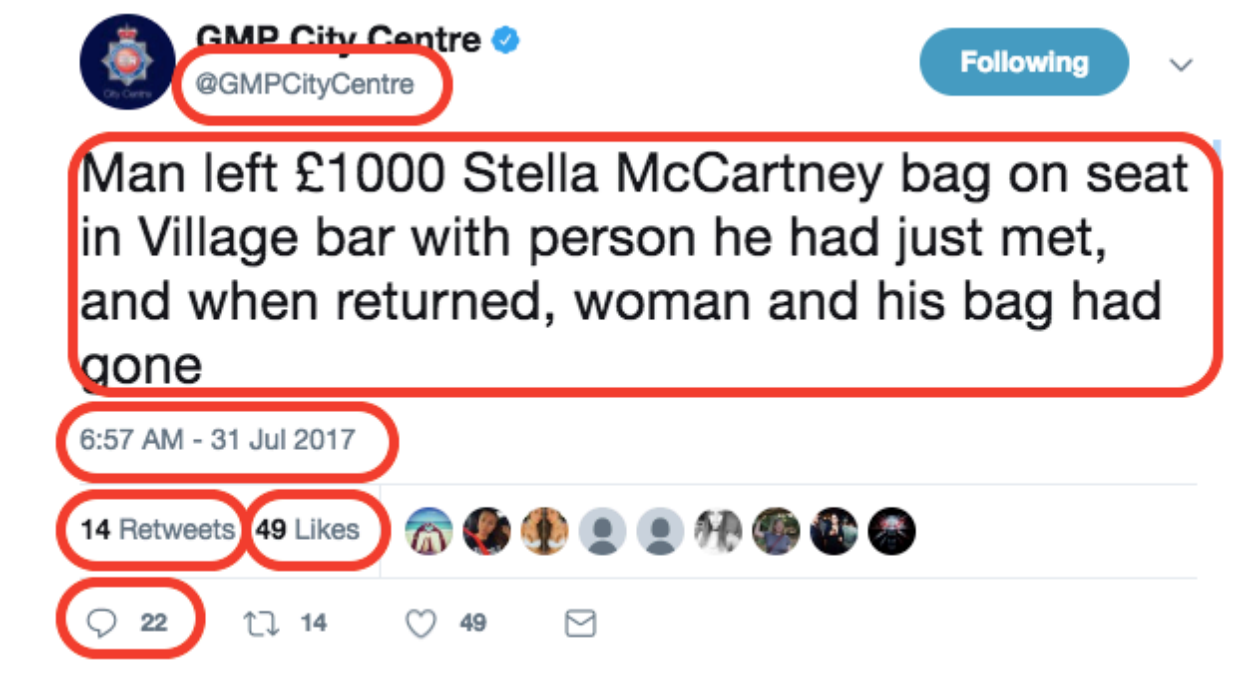So when you enter these values, your data will look like this:

Figure 1.4: Tweet1