QGIS for organisers

1. Intros
2. What we'll cover
    a. What QGIS is and the foundational concepts
    b. What kinds of data are useful to organisers and where to find them
    c. How to make a map of an electorate which shows polling booths and margins, as well as ABS demographic data for targeting
    d. (Maybe) a quick-and-dirty rough method for finding walkable blocks
3. What we won't
    a. Handling raster data in any detail
    b. A robust solution to the problem of finding walkable blocks
4. What QGIS is
    a. QGIS is a free, open-source, GIS, or Geographical Information System
        i. From wikipedia: "A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present spatial or geographic data. GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations"
    b. Get it here: https://qgis.org/en/site/forusers/download.html
    c. Which version?
        i. The new version 3 is faster and smoother to use, but it doesn't yet have all the plugins version 2 has. For most things version 3 will be easier to use.
    d. When would you use it over google maps?
        i. QGIS is much more powerful and flexible than something like google maps, when it comes to analysing and presenting data. But it's harder for anyone else to interact with what you produce beyond looking at it. You can make and share good maps, and if that's all you need to do, QGIS is good. If you need something interactive, where e.g. someone can turn layers on or off etc, google maps is better.
        ii. You can also make a map in QGIS then export it as a kml file and import it into Google Maps
5. Basic concepts
    a. Layers
        i. QGIS handles and displays data in layers, similar to e.g. how Photoshop or Illustrator does it. Each layer can be displayed or not, or be transparent to some degree, and we have lots of choices about how it's represented - colours and density and labels etc
    b. Projections

i. All maps and spatial data involve some kind of **projection,** or **CRS (coordinate reference system)** - some way of translating points on a sphere into points on a 2-D map or screen.
ii. QGIS can translate between these on the fly, so a lot of the time we can ignore them. But we need to pay attention when we import and save, because otherwise your data can end up on the other side of the world or in the middle of the ocean off the coast of Africa
iii. There are only a small handful I've had to deal with
    1. GDA94 - lots of x-y latitude-longitude pairs are in this format
    2. 3857 aka WGS84 pseudo mercator - Open Street Maps uses this

c. Types of geometric data
    i. **Points** have coordinates (Street addresses can be represented this way)
    ii. **Lines** are defined by at least two points (streets and roads can be represented this way)
    iii. **Polygons** are lines where the first and last point are the same (property boundaries, national boundaries, electorate boundaries etc)
d. Data QGIS can handle
    i. Two main types, raster and vector
    ii. Raster
        1. E.g. images. Raster data is a rectangular grid of dots. GIS systems deal with this a lot for satellite images etc. I've only used it when there's been a map I want to interrogate and the vector data isn't available, e.g. when the Age published a map of the new electoral boundaries and I wanted to see how they overlapped with the old ones.
        2. Raster filetypes: jpegs, tiffs etc
    iii. Vector - this is what I use the most
        1. Points, lines, polygons as described above
        2. Filetypes
            a. Shapefile
            b. CSVs can be treated as vector data if they have coordinate data in them
    iv. Other data
        1. You can import CSVs or databases which **don't** have any geometric data in them, and then attach that data to spatial data so you can map it.
        2. You can download database data, and connect to online databases
        3. You can connect to web-based mapping services, including google maps and open street maps
e. Plugins and tools
    i. Like a lot of platforms, QGIS lets third parties write 'plugins' which extend its capabilities, just as e.g. apps extend the capabilities of your phone

        ii. There's also a Toolbox with hundreds of processes in it you can use
6. What kinds of data are useful to organisers, and where are they?
    a. Electoral boundaries - polygon vector data
        i. Federal are here: https://aec.gov.au/Electorates/gis/gis_datadownload.htm
            1. There's multiple data types - afaik QGIS can handle them all
        ii. I assume state electoral commissions do the same, but haven't dealt much with state boundaries
    b. Electoral results
        i. This is available as CSV files from the AEC or state-based commissions. We can join it to spatial data to map it
    c. ABS data
        i. ABS have three kinds of data
            1. Boundaries / shapefiles
            2. Statistical data
            3. Geopackages, which combine the two
        ii. Boundaries/shapefiles are mostly here: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202016?OpenDocument
            1. Mesh blocks
                a. Mesh blocks are the smallest spatial unit the ABS has. Everything else (SA1s, etc) is built from them
                b. They're designed to have 30-60 households
                c. This is what Neighbourly uses to create walklists
                d. Most ABS data isn't publicly available at this scale, though orgs can enter into agreements with the ABS to access some data at this scale
            2. Statistical areas
                a. SA1s (statistical area 1) are the smallest unit we can get most demographic data from
            3. Postal areas
                a. For a lot of ACF supporters, postcode is the most we know about their location, so they're useful
                b. **Postcode** spatial data is commodified and v expensive
                c. Postal **areas** are a free version the ABS built out of mesh blocks, which are 99% identical to postcode boundaries
    d. There's lots of useful stuff available in some state-based data portals - VIC and QLD are pretty accessible. NSW is not. Not sure about other states.
        i. VIC spatial data: https://www.data.vic.gov.au/data/group/spatial-data
        ii. QLD spatial data: http://qldspatial.information.qld.gov.au/catalogue/custom/index.page
        iii. NSW (more confusing) : https://sdi.nsw.gov.au/nswsdi/catalog/main/home.page

7. Keeping data organised
    a. You very quickly can end up with a **lot** of data getting very messy. I am messy by nature so have learned the hard way to have a better folder structure.
    b. Currently I do this:
        i. Project folder
        ii. Data folder
            1. National level data
            2. State level data
                a. NSW data
                b. VIC data
                c. Etc
            3. Electorate-level data
                a. Macnamara data
                b. Chisholm data
                c. Etc
    c. Also shapefiles - a very common kind of spatial data - are directories of multiple files and should be saved as such into their own folders. They're gradually being replaced with single-file data types, but whenever you have shapefile, save it into its own subfolder
8. There are three big chunks of tasks with QGIS
    a. Finding and importing data. Skills/challenges here are
        i. Knowing where to find it,
        ii. Dealing with formats,
        iii. Saving it in a sensible way
        iv. Setting the right CRS
    b. Styling it
        i. Skills/challenges here are more about design - what colours work well together, how transparent to make things etc
    c. Analysis and data manipulation
        i. Here it's often about finding the right plugin and trial and error
9. Questions?
10. **Example 1: get a basemap**
    a. All the above is great, but we need a base layer to make sense of it - something which shows place names, roads etc. This is called a **basemap.**The easiest way to do this is use a web based map like Open Street Map or Google Maps
    b. Create a new project
        i. Use CRS 3857
    c. Download the plugin QuickMapServices
        i. >Plugins>Manage and install plugins
        ii. Unlike a lot of other platforms, most QGIS plugins populate other menus rather than being accessed through the plugin menu
    d. Add OSM as a basemap
        i. >Web>QuickMapServices>OSM>Osm Standard

e. Grab and zoom to your location of interest

11. Example 2: federal electorate boundaries / adding and styling a shapefile

a. Up-to-date federal electorate boundary files are always here: https://aec.gov.au/Electorates/gis/gis_datadownload.htm
b. (proposed boundary changes are somewhere else, until the changes are approved)
c. Download
d. Add vector layer
e. It covers the whole map, so we need to style it! Right click the layer in the layers panel and select >Properties>Symbology
f. There's lots of options. You can see up the top we can style areas according to characteristics, but for the moment we'll keep Simple fill
g. But click on Simple fill and we'll change some of its parameters
    i. Fill colour - transparent
    ii. Stroke colour - blue
    iii. Stroke width - bump it up a bit
h. Let's add a label >Properties>Labels
    i. >Label with>Sortname
    ii. Make it bold
    iii. And blue
i. It'll also help to filter the electorates so we're only looking at the one we want
j. Right click the layer in the layers panel and select Filter
    i. Double click Sortname from the list at the top
    ii. Click the = button
    iii. Type 'Wentworth', and apply

12. Example 3: polling booth locations / adding and styling CSV point data

a. Expected polling booth locations for the next federal election are here: https://www.aec.gov.au/About_AEC/cea-notices/election-pp.htm
    i. These lists get updated periodically, so worth checking back closer to the election too
b. Download and open the doc
c. We need to remove any commas! Find and replace all commas with spaces. (CSVs use commas to separate columns, so commas within a cell confuse QGIS.)
d. Save
e. In QGIS, >Layer>Add layer>Add delimited text layer
f. CRS = 4326 - WGS 84
g. If it's worked, you should have a lot of dots on your map!
h. You can right-click on the layer in the layers panel and choose how to style it
i. We can also filter
j. Questions?

13. Example 4: add 2PP data to polling booths / joining non-spatial data to spatial data

a. Finding and adding 2PP data

i. So we have polling booth locations, but that tells us nothing about the margin. We can download AEC data and work out the margin, but that doesn't have latitude/longitude info in it, so we need to join it to the data we just imported so we can map it
ii. Here's a stack of results we could use: https://results.aec.gov.au/20499/Website/HouseDownloadsMenu-20499-Csv.htm
iii. Let's download **Two party preferred by polling place,** and open it
iv. I always remove the top row - it just confuses QGIS

b. *Optional - add margin info*
   i. *Add a new column called 'lab-lib'. This will show the ALP 2PP percentage minus the Coalition 2PP percentage*
   ii. *In the second-top in that column cell type* ***=sum(i2-g2)***
   iii. *Double click the bottom right corner of the cell - this should populate the whole column with this formula*
   iv. Save
   v. Add it by clicking **Add delimited text layer** in QGIS. This data has no coordinates in it, so click **No geometry (attribute only table).** This kind of data is attribute data, not spatial. You'll notice nothing shows up on the map when we add it, because there's no spatial data to display.

c. Joining electoral data to location data
   i. The next thing is where a lot of the power of QGIS is - joining non-spatial data to spatial data so we can map it
   ii. Right click on the polling place spatial data row in the layers panel, and select **Properties**, then **Joins**
   iii. First we need to tell it what layer we're joining to this one - the join layer
   iv. What we need to connect them is to find a column which is in both our spatial data and the data we want to join. There are a few candidates, but let's use the polling place ID - this is a number ID for each polling place. It has a slightly different name in each of our CSV files, but that doesn't matter
   v. So in *Join field*, select **PollingPlaceID**
   vi. In *Target field*, select **PPId**
   vii. Click OK

d. Label by vote
   i. Now if we want to label the polling places with, e.g. the Labor percentage, we can

14. Example 5: adding and styling census data

a. Census data is available in two forms: as CSVs of attributes with no location data (which we can join to boundaries as above), and as Geopackages, which combine attributes with location data so we don't need to join anything. I've found Geopackages a bit unwieldy, but it could be my version of QGIS.
b. Go to https://datapacks.censusdata.abs.gov.au/geopackages/ and download the selected person characteristics geopackage A for NSW
c. Add as a vector layer

i. Choose the SA1 layer, e.g. census2016_spca_nsw_short

d. We want to know the number of people aged 18-35 as a percentage of the total population
e. >Properties>Fields
    i. Open the field calculator by clicking on the abacus at the top
    ii. The panel on the left is for putting together formulae, a little similar to the kind of thing we did earlier in excel. The panel on the right gives us the bits we can build those formulae from
    iii. On the right, the subsection **Fields and Values** gives us access to the attribute data. Double clicking on values adds them to the left panel.
    iv. Since there's no field for percentage of 18-35 year olds, we need to build one
    v. Because of the way the data is broken down, we want to end up with
        1. **("Age_20_24_yr_P" + "Age_25_34_yr_P" + "Age_yr_19_P" + "Age_yr_18_P") / "Tot_P_P" * 100**
        2. Which means
            a. The number of 20-24 year olds in this statistical area
            b. Plus the number of 25-34 year olds
            c. Plus the number of 19 year olds
            d. Plus the number of 18 year olds
            e. ..all divided by the total number of people
            f. And multiplied by 100 to make a percentage
    vi.
f. Right click on the layer in the layer panel, click >Properties>Symbology and select **Graduated.** This lets us display something differently (either by size or colour) depending on where it sits on some scale
g. We need to choose how many categories - I usually start with 5 and adjust from there
h. For method I usually use Natural Breaks

15. If we have time: approximating knockable blocks

a. Load an address file. These are hard to find for NSW!
    i. Filter out units and apartments
b. Load mesh blocks
c. Make a subset for the electorate
d. >Processing>Count points in polygon
    i. Points layer is the address file
    ii. Polygons are the mesh blocks
e. It will return a set of polygons which are basically a copy of the mesh block file with an added count column
f. Deselect the old mesh blocks layer
g. Style the new mesh blocks layer so the fill is transparent and the outlines are a distinct colour
h. Filter the mesh blocks so the count value is >19 (or whatever cutoff you want)