Spatial data with R: an introduction

Wed 24th April
9am – 10:45am

Dr Nick Bearman & Dr Robin Lovelace

@nickbearmanuk

Wifi: ‘eduroam’ or ‘Wifi Guest’

Install R https://cran.r-project.org/
Install RStudio https://www.rstudio.com/

```
install.packages(“sf”)
install.packages(“tmap”)
```

If it is not working, ask!

http://bit.ly/GISRUK-2019

Geospatial Training Solutions

Geographical Information Science Research - UK

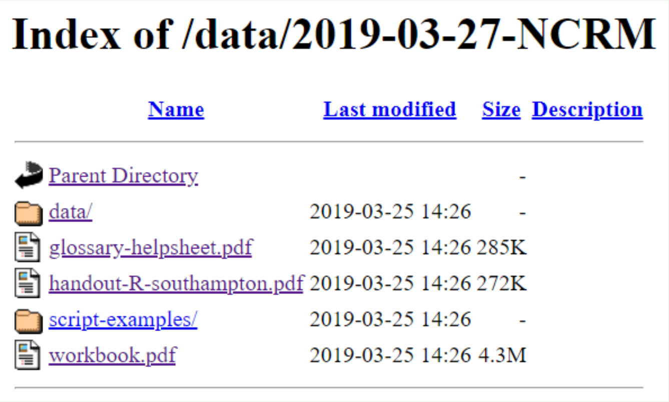
GISRUK

Housekeeping

- Log on!
- Toilets
- Fire Alarm

- Presentations, handouts and data online

http://bit.ly/GISRUK-2019

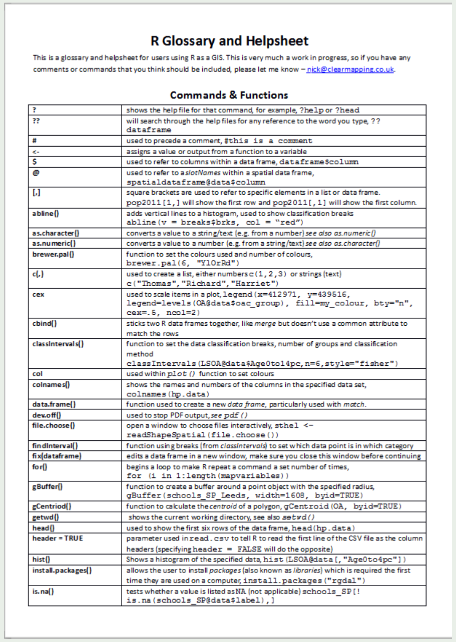


Index of /data/2019-03-27-NCRM

Name	Last modified	Size	Description
Parent Directory		-	
data/	2019-03-25 14:26	-	
glossary-helpsheet.pdf	2019-03-25 14:26	285K	
handout-R-southampton.pdf	2019-03-25 14:26	272K	
script-examples/	2019-03-25 14:26	-	
workbook.pdf	2019-03-25 14:26	4.3M	

R as a GIS

- Command line driven, rather than GUI
- Disadvantages
  - Remembering commands
    - *glossary*
  - Steeper learning curve

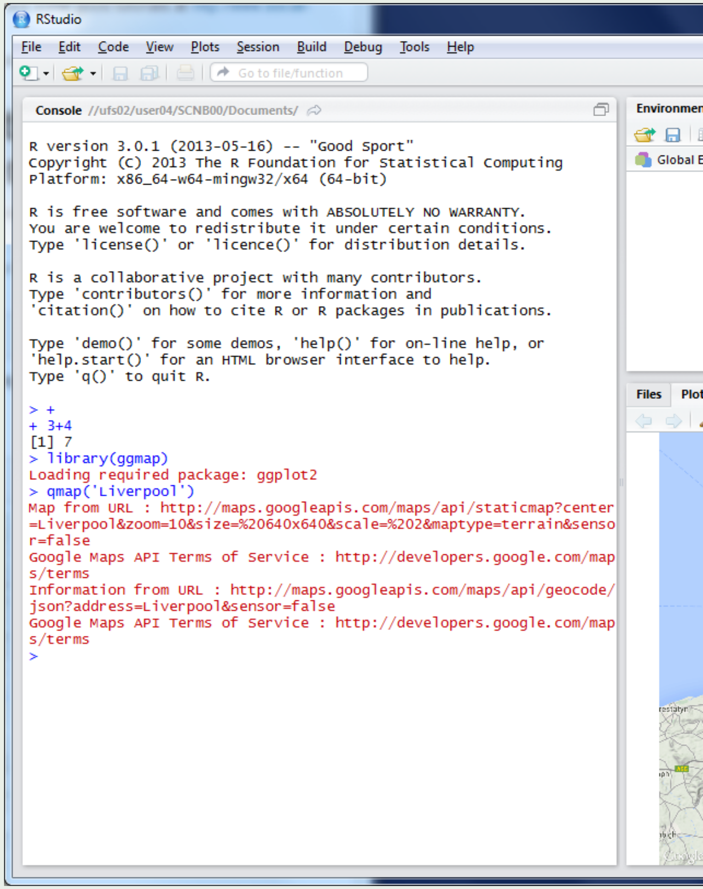


R Glossary and Helpsheet

This is a glossary and helpsheet for users using R as a GIS. This is very much a work in progress, so if you have any comments or commands that you think should be included, please let me know – nick@clearmapping.co.uk.

Commands & Functions

?	shows the help file for that command, for example, `?help` or `?head`
??	will search through the help files for any reference to the word you type, `?? dataframe`
#	used to precede a comment, `#this is a comment`
<-	assigns a value or output from a function to a variable
$	used to refer to columns within a data frame, `dataframe$column`
@	used to refer to a *slotNames* within a spatial data frame, `spatialdataframe@data$column`
[,]	square brackets are used to refer to specific elements in a list or data frame. `pop2011[1,]` will show the first row and `pop2011[,1]` will show the first column.
abline()	adds vertical lines to a histogram, used to show classification breaks `abline(v = breaks$brks, col = “red”)`
as.character()	converts a value to a string/text (e.g. from a number) *see also as.numeric()*
as.numeric()	converts a value to a number (e.g. from a string/text) *see also as.character()*
brewer.pal()	function to set the colours used and number of colours, `brewer.pal(6, "YlOrRd")`
c(,)	used to create a list, either numbers `c(1,2,3)` or strings (text) `c("Thomas","Richard","Harriet")`
cex	used to scale items in a plot, `legend(x=412971, y=439516, legend=levels(OA@data$oac_group), fill=my_colour, bty="n", cex=.5, ncol=2)`
cbind()	sticks two R data frames together, like *merge* but doesn't use a common attribute to match the rows
classIntervals()	function to set the data classification breaks, number of groups and classification method `classIntervals(LSOA@data$Age0to14pc,n=6,style="fisher")`
col	used within *`plot()`* function to set colours
colnames()	shows the names and numbers of the columns in the specified data set, `colnames(hp.data)`
data.frame()	function used to create a new *data frame*, particularly used with *match*.
dev.off()	used to stop PDF output, *see `pdf()`*
file.choose()	open a window to choose files interactively, `sthel <- readShapeSpatial(file.choose())`
findInterval()	function using breaks (from *classIntervals*) to set which data point is in which category
fix(dataframe)	edits a data frame in a new window, make sure you close this window before continuing
for()	begins a loop to make R repeat a command a set number of times, `for (i in 1:length(mapvariables))`
gBuffer()	function to create a buffer around a point object with the specified radius, `gBuffer(schools_SP_Leeds, width=1608, byid=TRUE)`
gCentriod()	function to calculate the *centroid* of a polygon, `gCentroid(OA, byid=TRUE)`
getwd()	shows the current working directory, see also *`setwd()`*
head()	used to show the first six rows of the data frame, `head(hp.data)`
header = TRUE	parameter used in `read.csv` to tell R to read the first line of the CSV file as the column headers (specifying `header = FALSE` will do the opposite)
hist()	Shows a histogram of the specified data, `hist(LSOA@data[,"Age0to4pc"])`
install.packages()	allows the user to install *packages* (also known as *libraries*) which is required the first time they are used on a computer, `install.packages("rgdal")`
is.na()	tests whether a value is listed as NA (not applicable) `schools_SP[!is.na(schools_SP@data$label),]`


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
Console //ufs02/user04/SCNB00/Documents/
Environmen
Global E
Files Plot

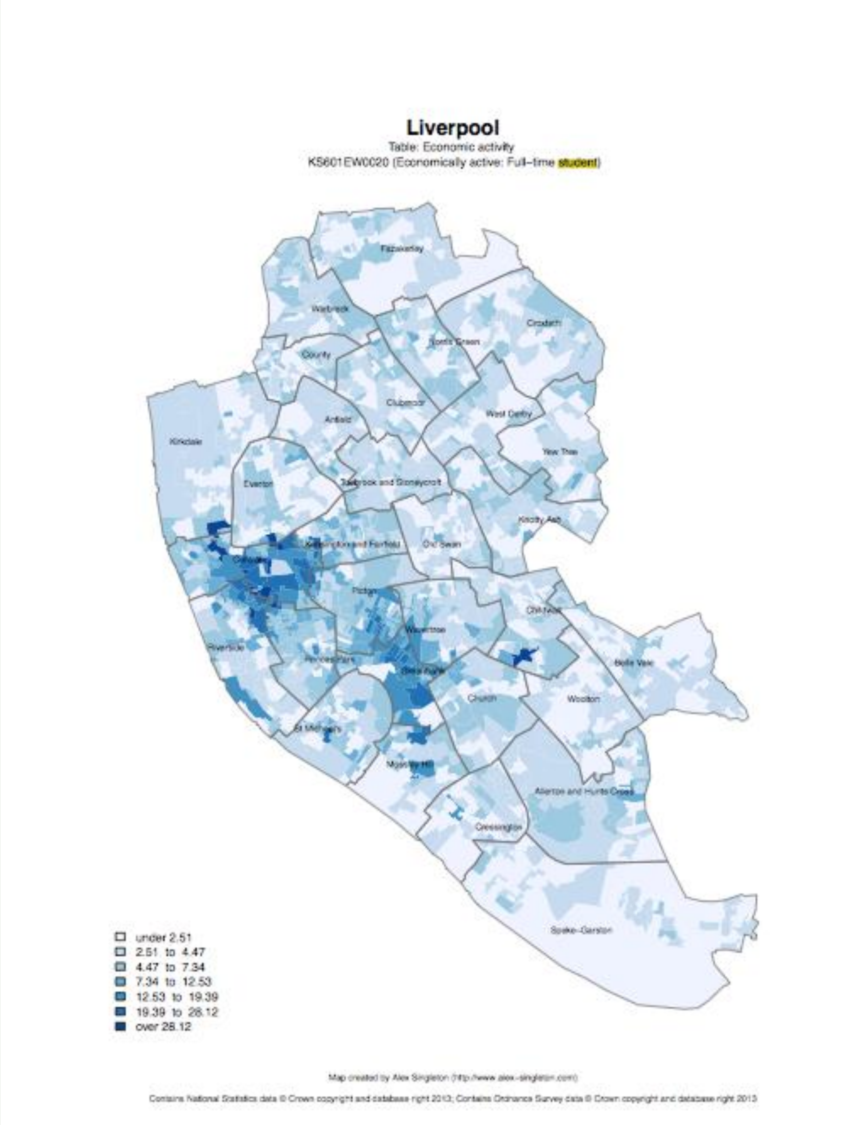
```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> +
+ 3+4
[1] 7
> library(ggmap)
Loading required package: ggplot2
> qmap('Liverpool')
Map from URL : http://maps.googleapis.com/maps/api/staticmap?center
=Liverpool&zoom=10&size=%20640x640&scale=%202&maptype=terrain&senso
r=false
Google Maps API Terms of Service : http://developers.google.com/map
s/terms
Information from URL : http://maps.googleapis.com/maps/api/geocode/
json?address=Liverpool&sensor=false
Google Maps API Terms of Service : http://developers.google.com/map
s/terms
>
```

R as a GIS

- Advantages
  - Easy to record what you did and repeat specific pieces of work
  - Lots of reproducible examples on the web
  - Easily scriptable.
  - 134,567 maps? Easy! (354 areas X ~392 variables)
  - 2011 Census Open Atlas

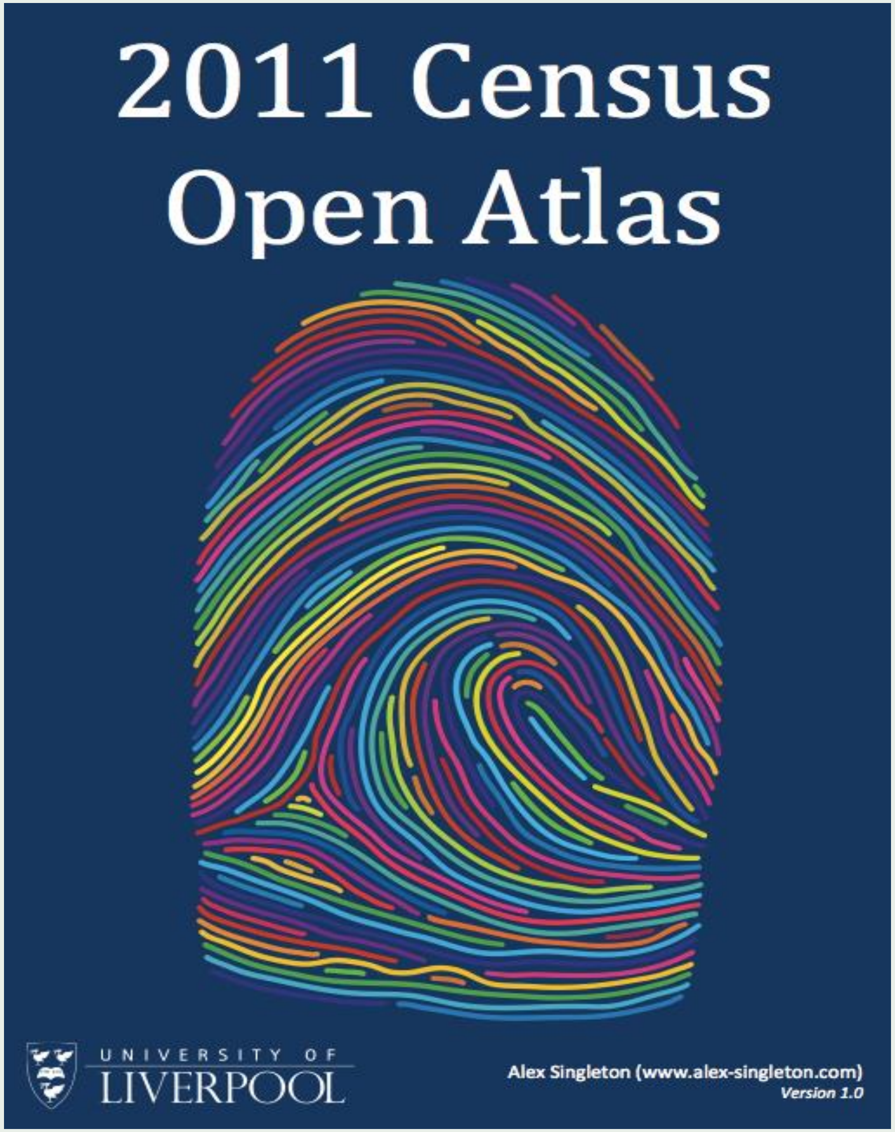
  - http://www.alex-singleton.com/r/2014/02/05/2011-census-open-atlas-project-version-two/

R in Action

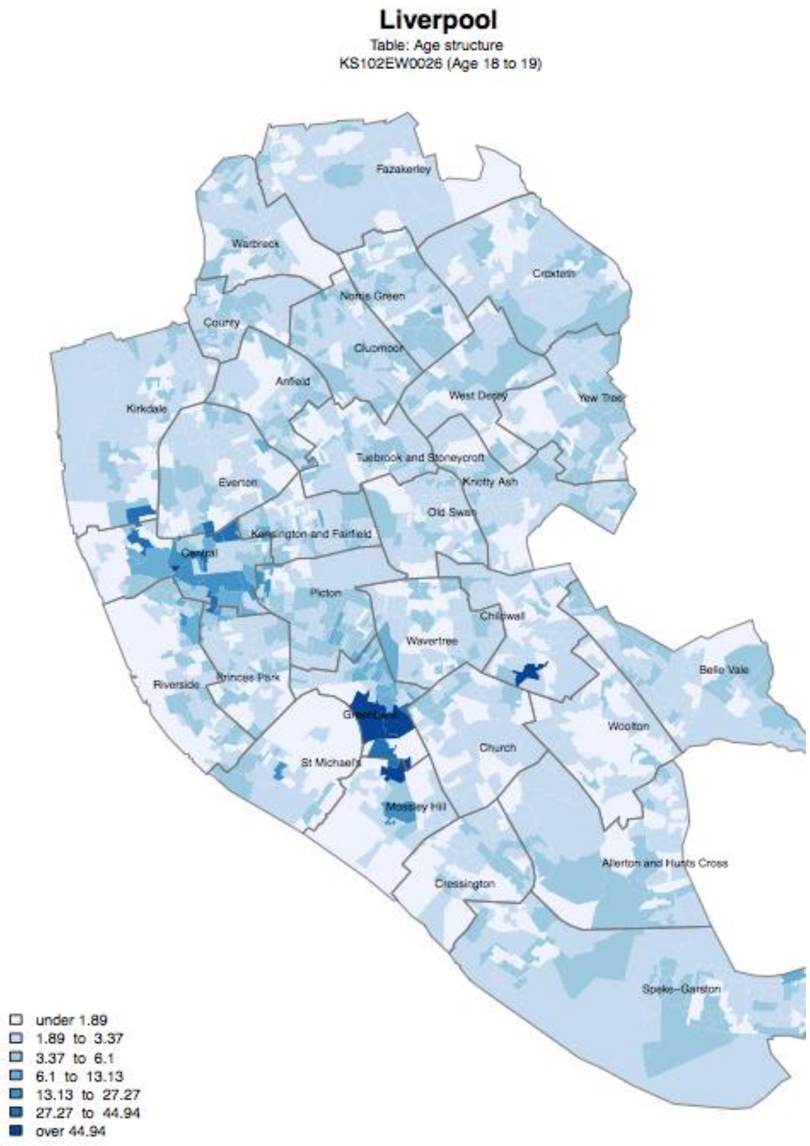


Map created by Alex Singleton (http://www.alex-singleton.com)

Contains National Statistics data © Crown copyright and database right 2013; Contains Ordnance Survey data © Crown copyright and database right 2013



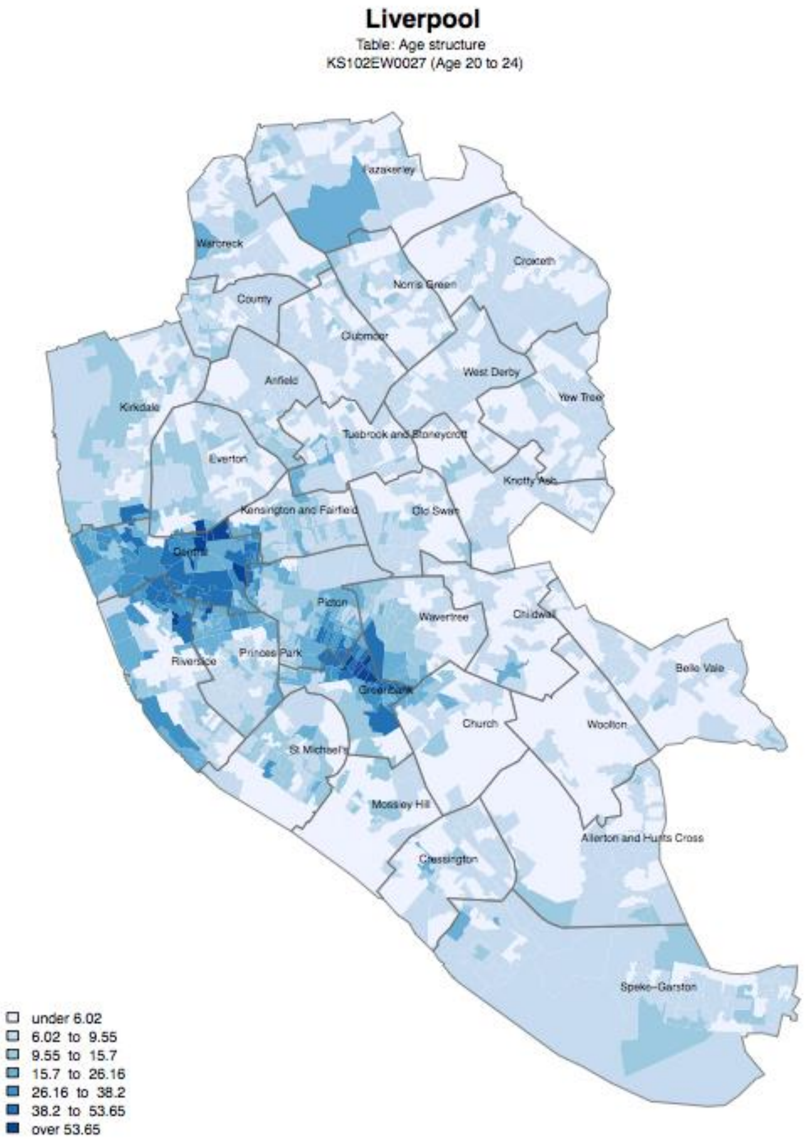
Economically active – Full-time students



Map created by Alex Singleton (http://www.alex–singleton.com)

Contains National Statistics data © Crown copyright and database right 2013; Contains Ordnance Survey data © Crown copyright and database right 2

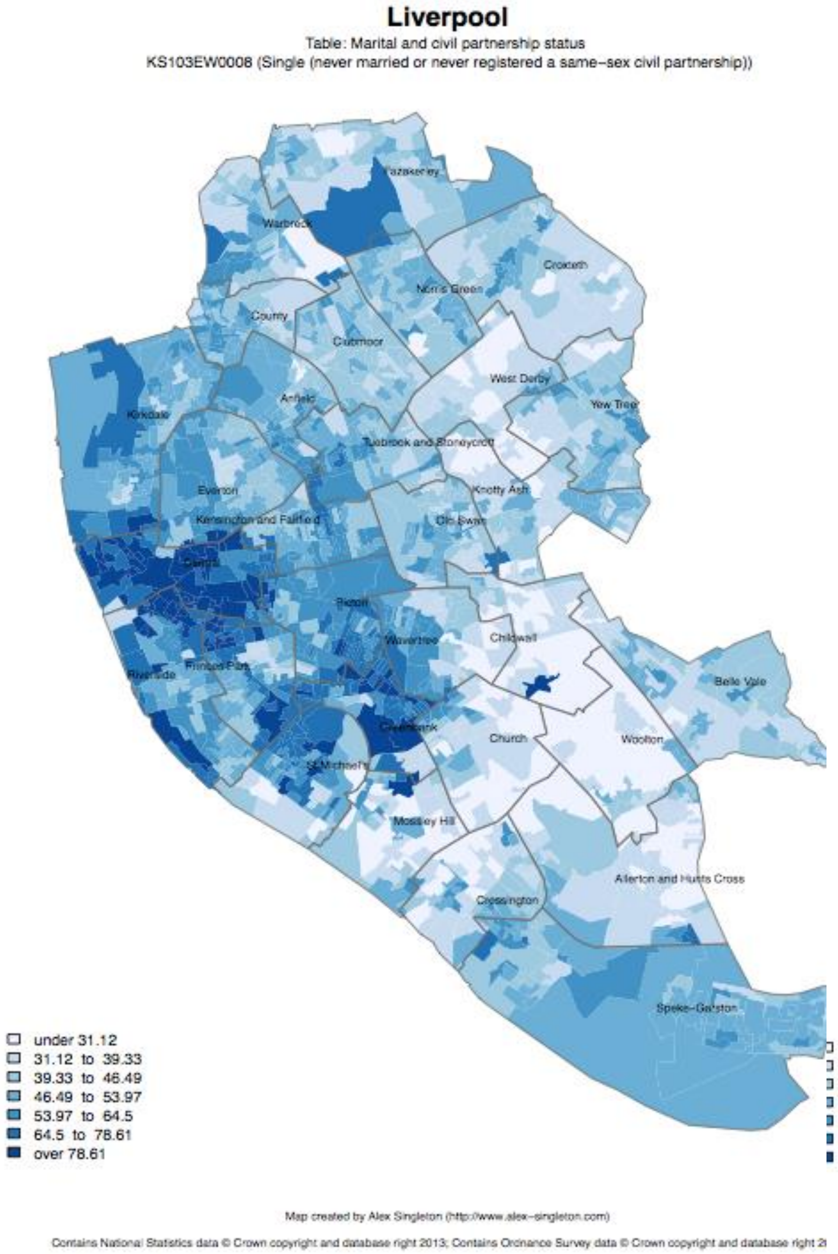
Age structure – 18 to 19



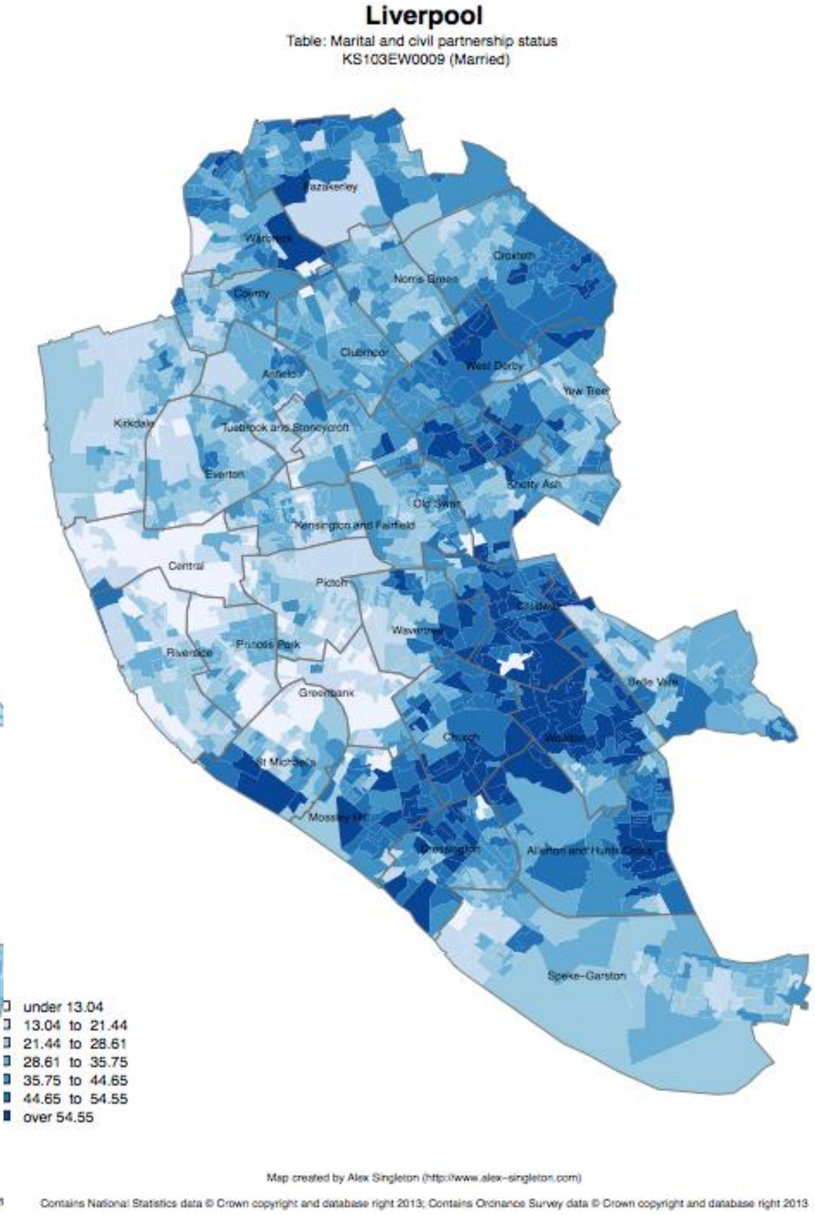
Map created by Alex Singleton (http://www.alex–singleton.com)

Contains National Statistics data © Crown copyright and database right 2013; Contains Ordnance Survey data © Crown copyright and database right 2013

Age structure – 20 to 24



Single

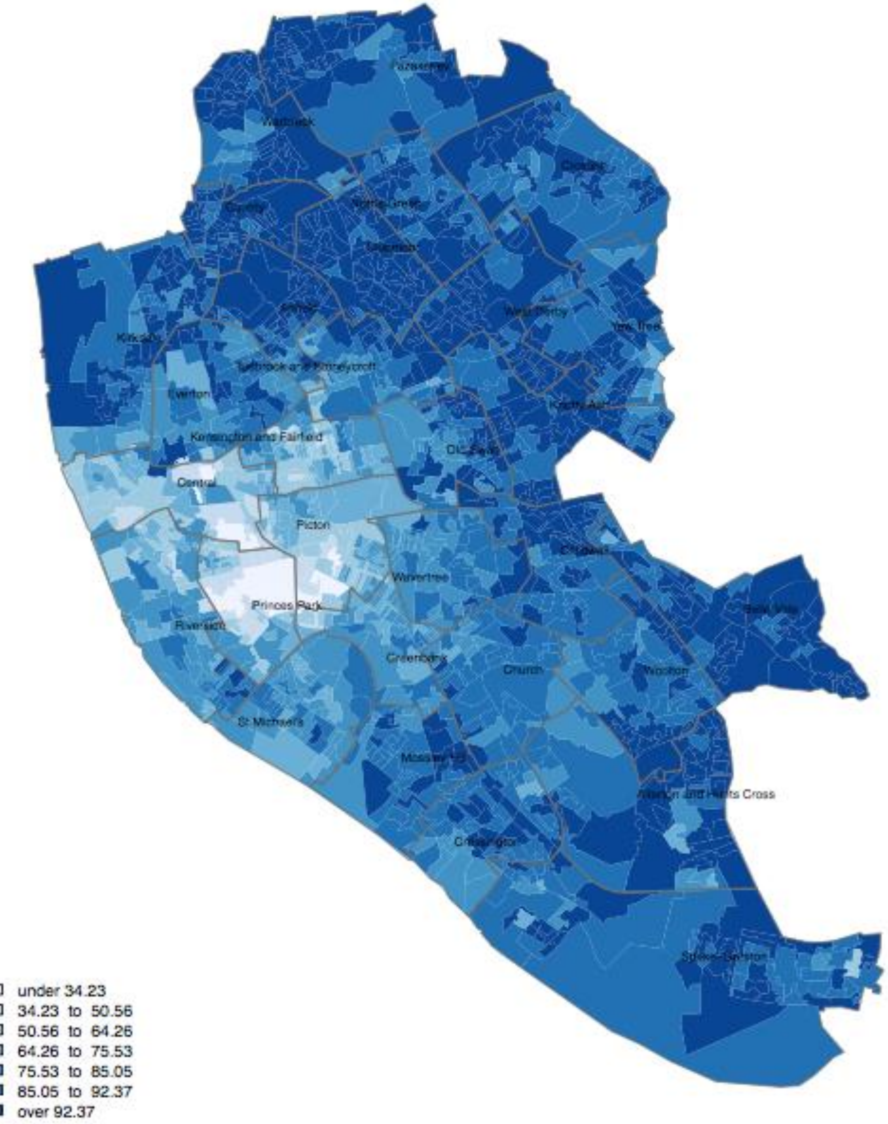


Married

Liverpool

Table: Ethnic group

KS201EW0020 (White: English/Welsh/Scottish/Northern Irish/British)



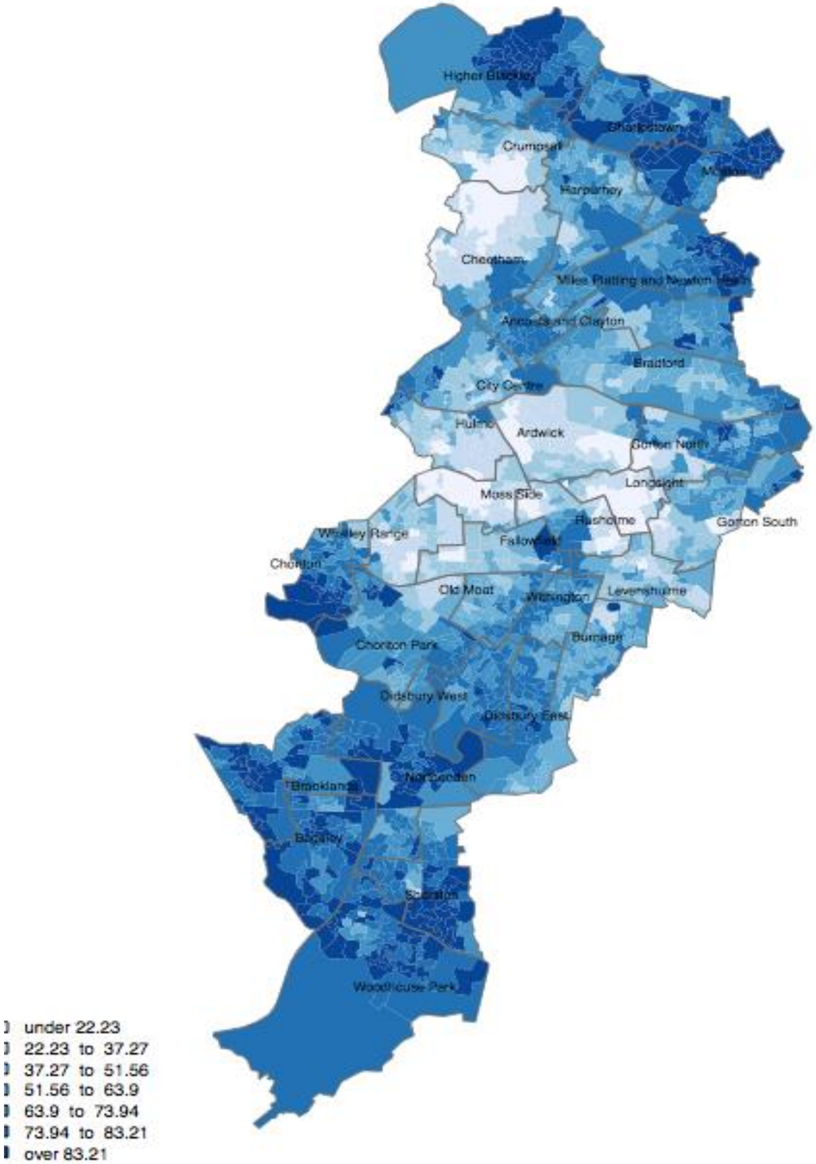
Map created by Alex Singleton (http://www.alex-singleton.com)

Contains National Statistics data © Crown copyright and database right 2013; Contains Ordnance Survey data © Crown copyright and database right 2013

Manchester

Table: Ethnic group

KS201EW0020 (White: English/Welsh/Scottish/Northern Irish/British)

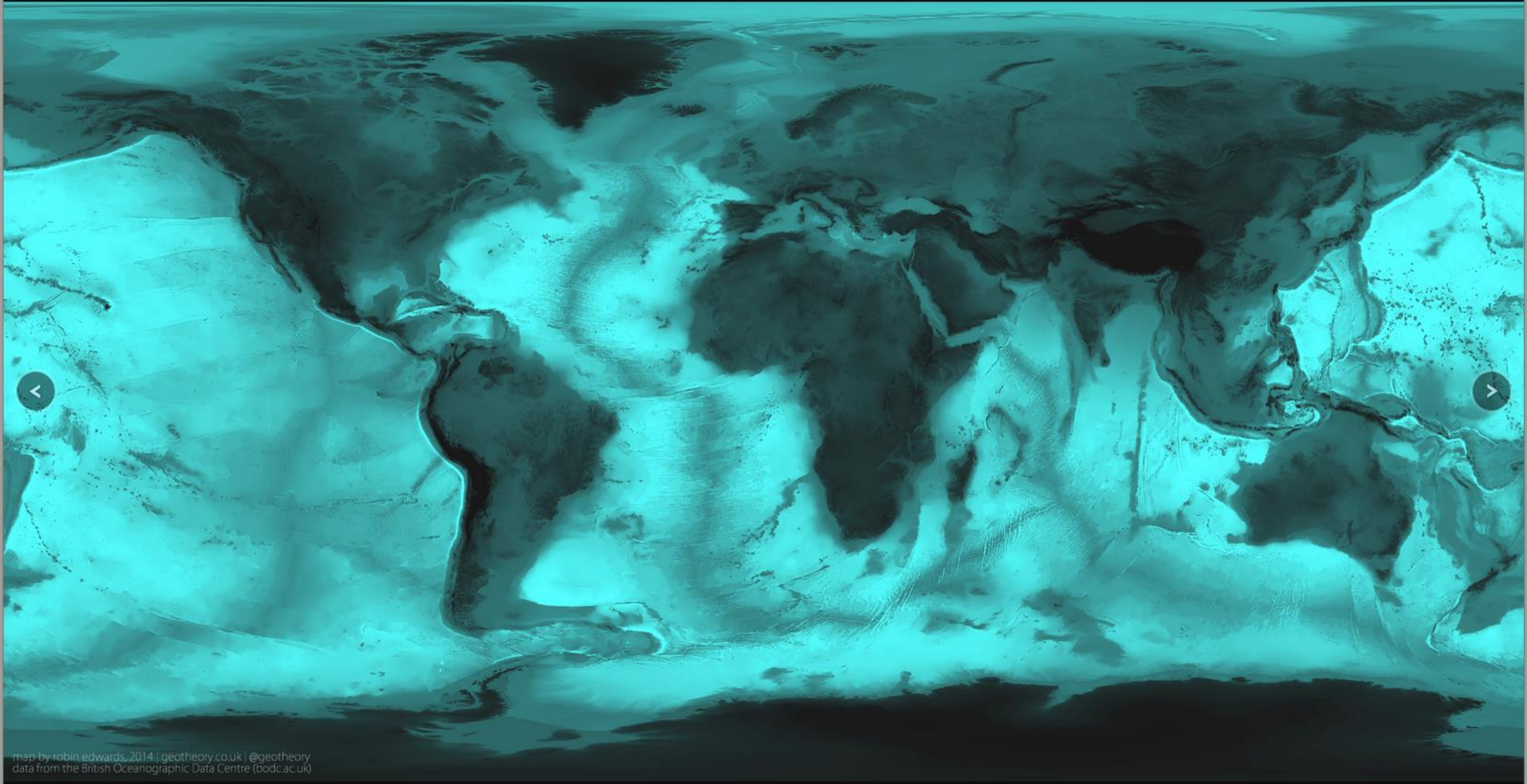


Map created by Alex Singleton (http://www.alex-singleton.com)

Contains National Statistics data © Crown copyright and database right 2013; Contains Ordnance Survey data © Crown copyright and database right 2

Ethnic group: white

Topography



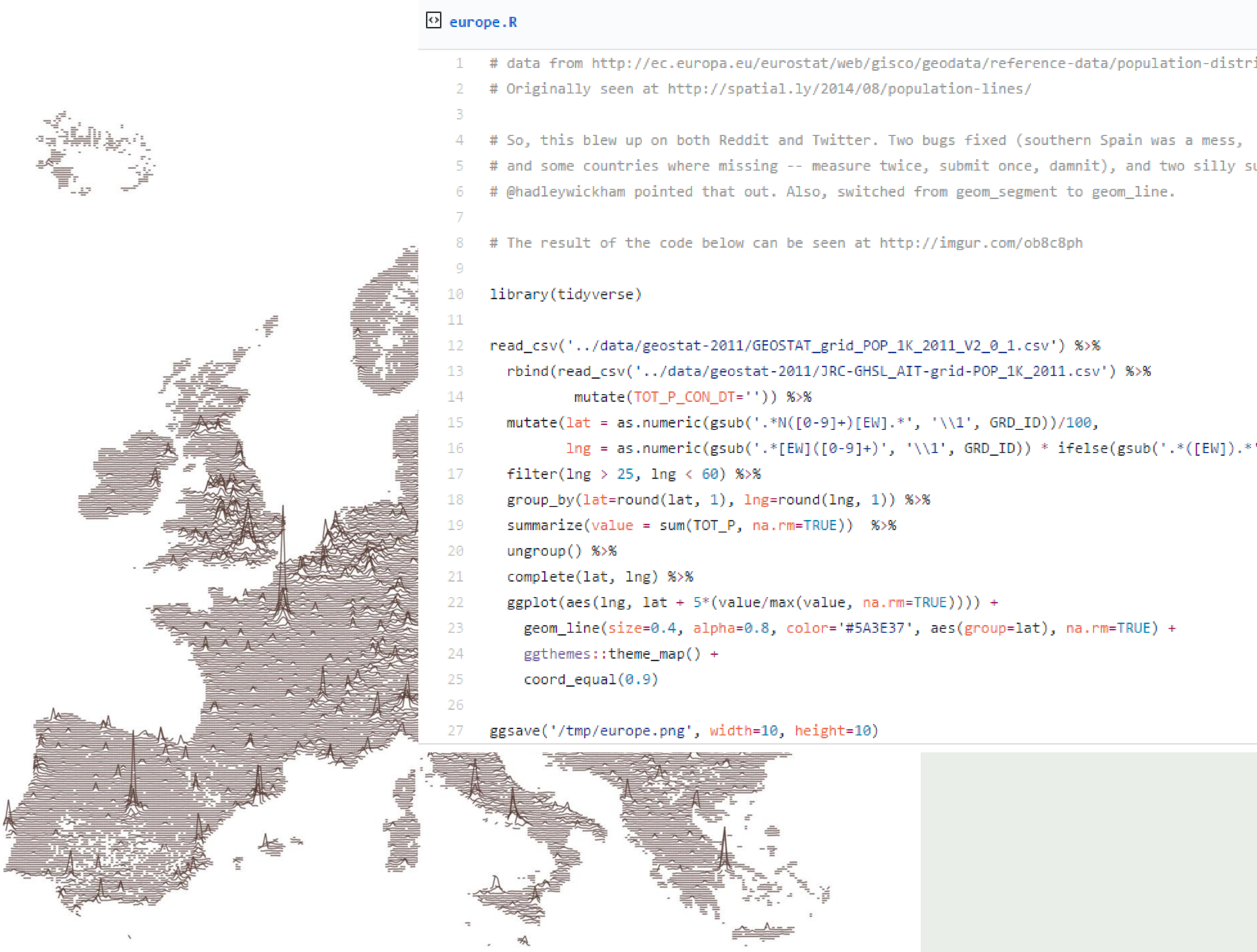
map by robin edwards, 2014 | geotheory.co.uk | @geotheory
data from the British Oceanographic Data Centre (bodc.ac.uk)

http://topography.geotheory.co.uk/

1 / 13 Global map of ocean depths and land elevations, with data mapped to colour (cyan to dark grey) by quantile.

topography.geotheory.co.uk/slides/world.png

https://gist.github.com/halhen/659780120accd82e043986c8b57deae



europe.R

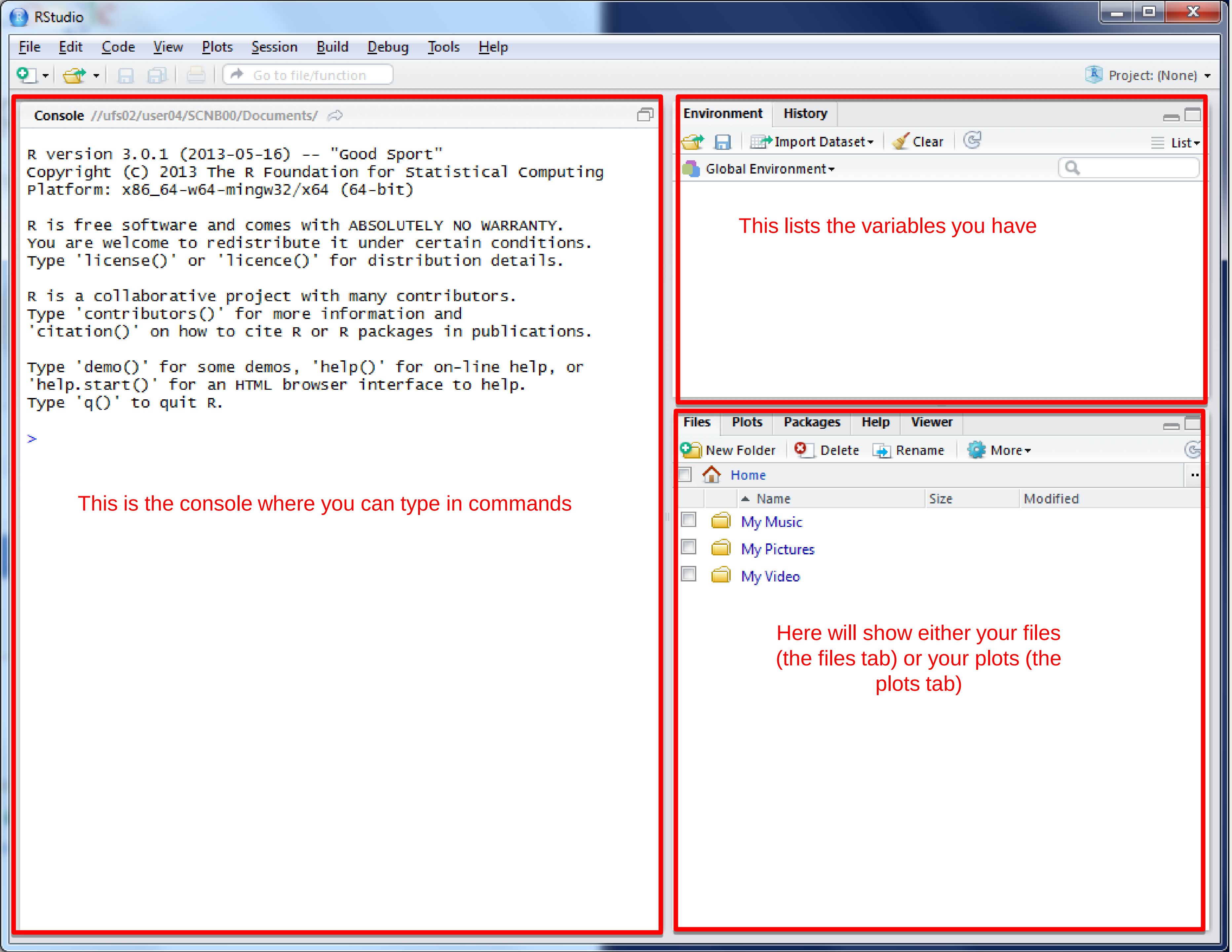
```r
# data from http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distri
# Originally seen at http://spatial.ly/2014/08/population-lines/

# So, this blew up on both Reddit and Twitter. Two bugs fixed (southern Spain was a mess,
# and some countries where missing -- measure twice, submit once, damnit), and two silly su
# @hadleywickham pointed that out. Also, switched from geom_segment to geom_line.

# The result of the code below can be seen at http://imgur.com/ob8c8ph

library(tidyverse)

read_csv('../data/geostat-2011/GEOSTAT_grid_POP_1K_2011_V2_0_1.csv') %>%
  rbind(read_csv('../data/geostat-2011/JRC-GHSL_AIT-grid-POP_1K_2011.csv') %>%
          mutate(TOT_P_CON_DT='')) %>%
  mutate(lat = as.numeric(gsub('.*N([0-9]+)[EW].*', '\\1', GRD_ID))/100,
         lng = as.numeric(gsub('.*[EW]([0-9]+)', '\\1', GRD_ID)) * ifelse(gsub('.*([EW]).*'
  filter(lng > 25, lng < 60) %>%
  group_by(lat=round(lat, 1), lng=round(lng, 1)) %>%
  summarize(value = sum(TOT_P, na.rm=TRUE))  %>%
  ungroup() %>%
  complete(lat, lng) %>%
  ggplot(aes(lng, lat + 5*(value/max(value, na.rm=TRUE)))) +
    geom_line(size=0.4, alpha=0.8, color='#5A3E37', aes(group=lat), na.rm=TRUE) +
    ggthemes::theme_map() +
    coord_equal(0.9)

ggsave('/tmp/europe.png', width=10, height=10)
```

https://gist.github.com/halhen/659780120accd82e043986c8b57deae0

Other GIS software?

- R is very different to ArcGIS, but can do many of the same operations
- R is free (as is QGIS)
- R is increasingly popular in academic sector
- *Try data in QGIS if you like*

All can be useful


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
Project: (None)
Console //ufs02/user04/SCNB00/Documents/
Environment History
Import Dataset Clear List
Global Environment
Files Plots Packages Help Viewer
New Folder Delete Rename More
Home
Name Size Modified
My Music
My Pictures
My Video

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

This lists the variables you have

This is the console where you can type in commands

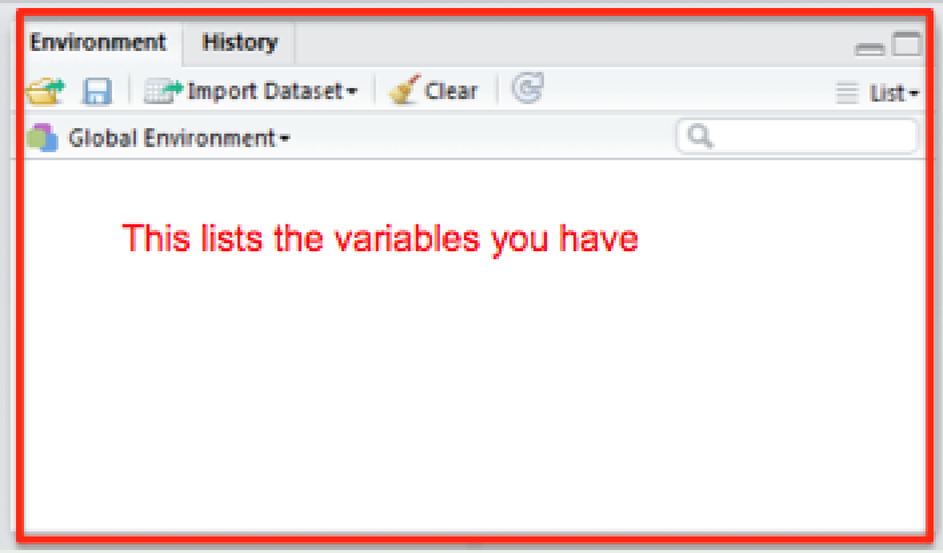
Here will show either your files (the files tab) or your plots (the plots tab)

Working Directory

- R uses a ‘working directory’ to store your files in
- You might have a different one for each project / piece of work
- e.g. M:\Documents\GIS

```
setwd(“M:/Documents/GIS”)
```

Variables



This lists the variables you have

R uses variables to store information – listed in your ‘workspace’ (top-right)

```
house.prices <- c(120,150,212,
99,199,299,159)
```

Variables and Assignments

```
house.prices
<-
c(120,150,212,99,199,299,159)
```

Variables and Assignments

```
house.prices
<-
c(120,150,212,99,199,299,159)
```

Variables and Assignments

```
house.prices
<-
c(120,150,212,99,199,299,159)
```

Variables and Assignment

```
sthelens
<-
st_read("sthelens.shp")
```

Variables and Assignment

```
sthelens
<-
st_read("sthelens.shp")
```

Variables and Assignment

```
sthelens
<-
st_read("sthelens.shp")
```

Variables and Assignment

```
sthelens
<-
st_read("sthelens.shp")
```

Variables and Assignment

```
sthelens
<-
st_read("sthelens.shp")
```

- Case sensitive

```
StHelens ≠ sthelens ≠ STHELENS
```

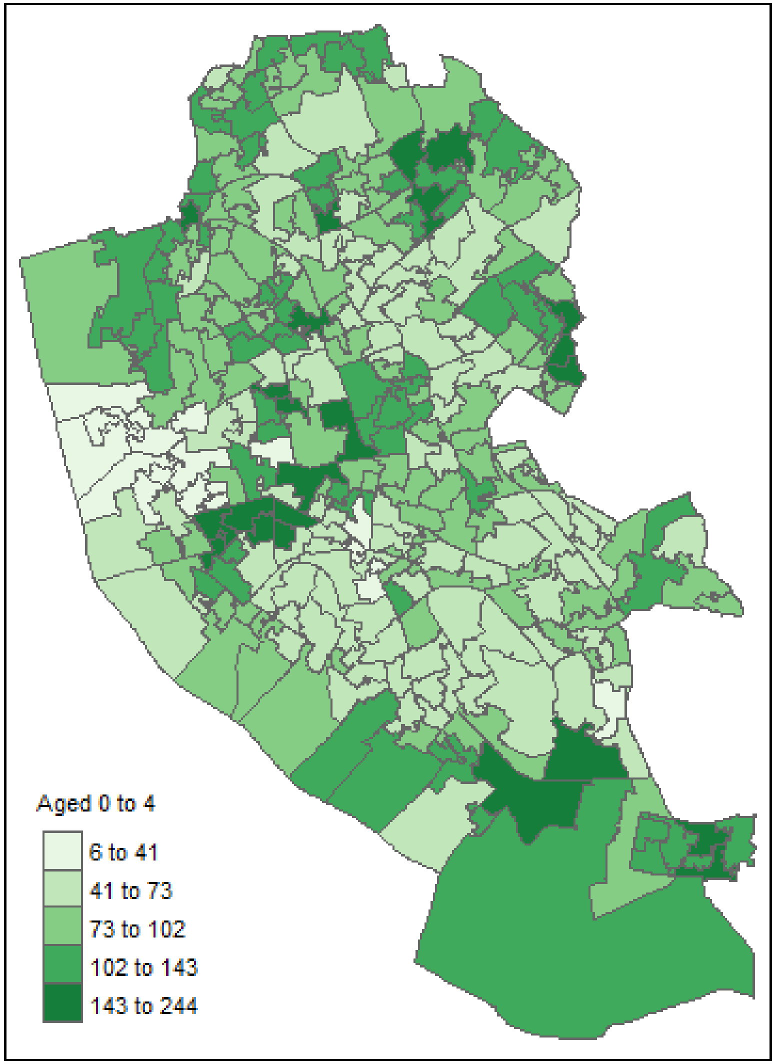
Data Formats

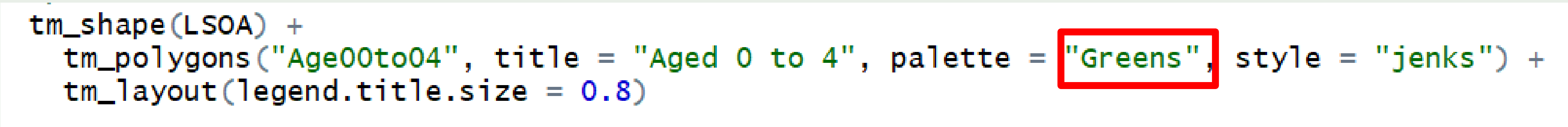
- Data frames are like a table or spreadsheet
- `dataframe[rows,cols]`
- `dataframe[1,]` first row
- `dataframe[,1]` first col

Practical Approach

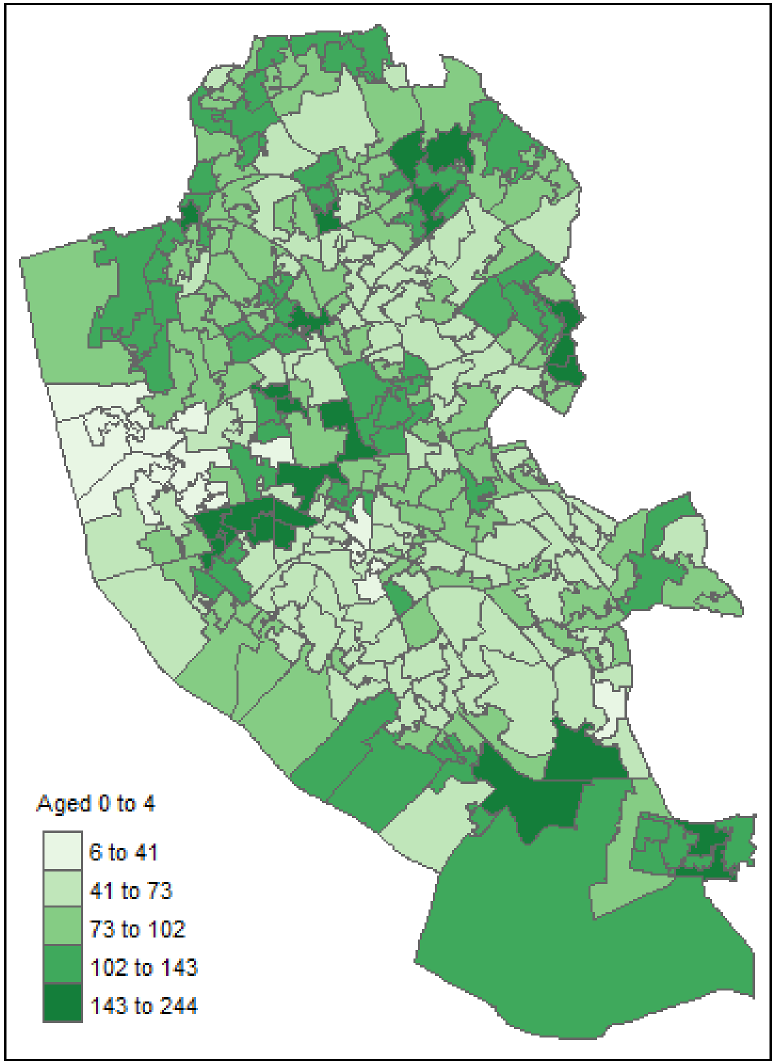
- (Primarily) self-led practical
- Good to explore the options
- Ask questions as we go through
- Try things
- Use the help `?command`
- You will need to install the libraries
    - `install.packages(“sf”)`
    - `library(sf)`

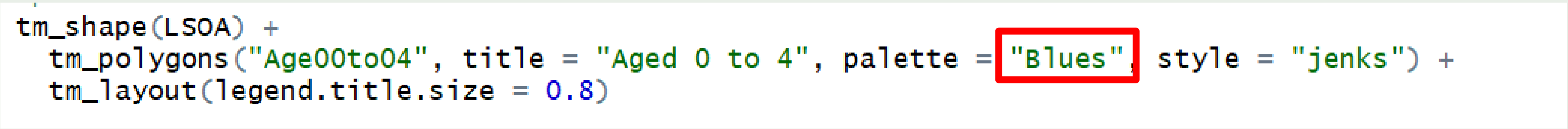
```r
tm_shape(LSOA) +
  tm_polygons("Age00to04", title = "Aged 0 to 4", palette = "Greens", style = "jenks") +
  tm_layout(legend.title.size = 0.8)
```



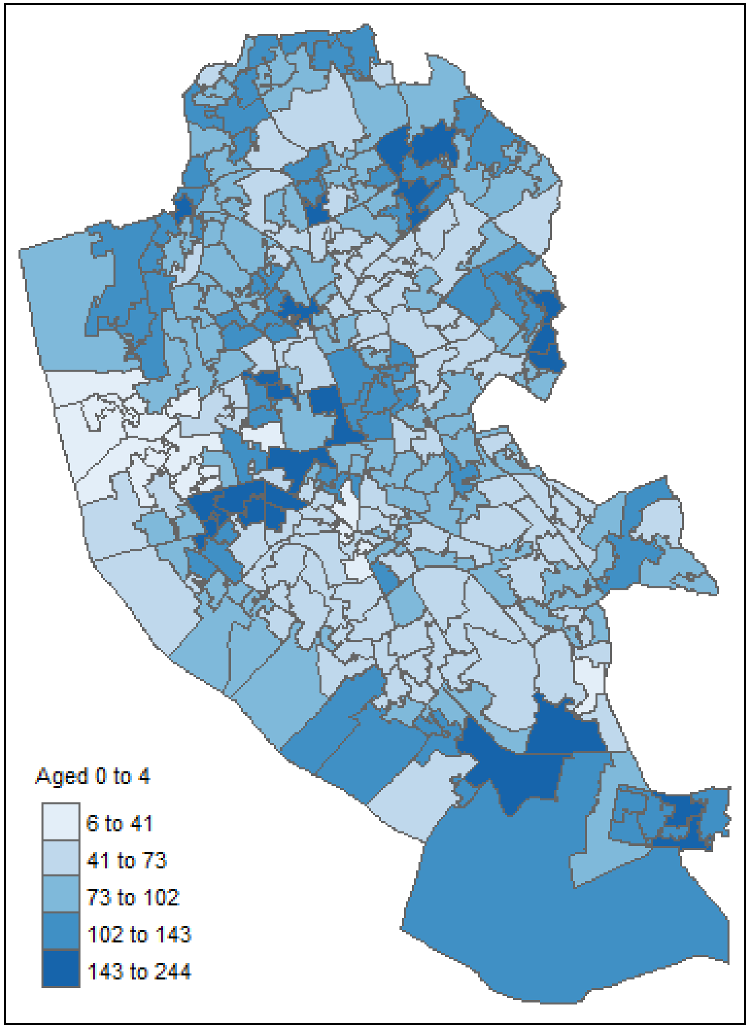


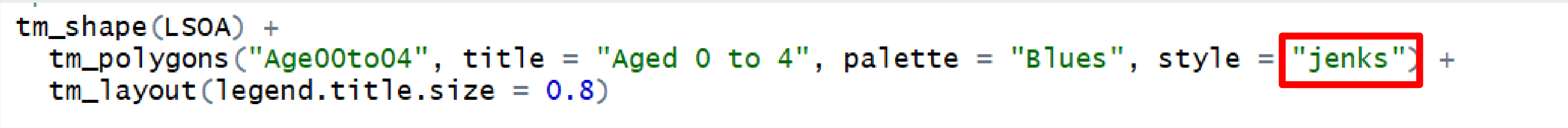
```
tm_shape(LSOA) +
  tm_polygons("Age00to04", title = "Aged 0 to 4", palette = "Greens", style = "jenks") +
  tm_layout(legend.title.size = 0.8)
```



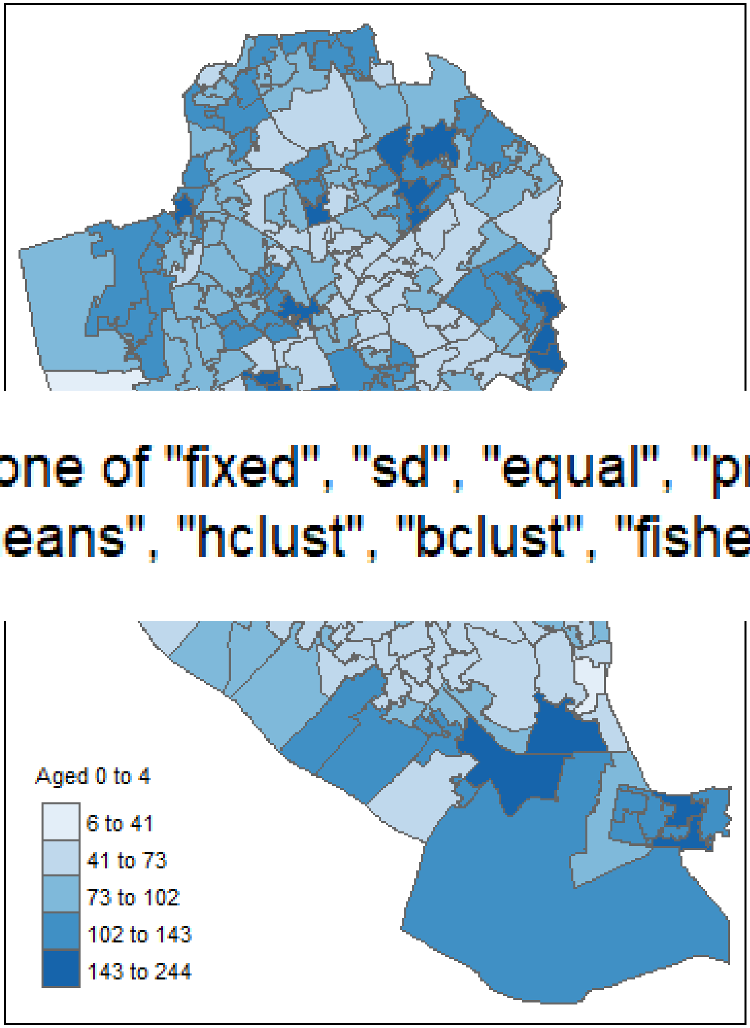


```
tm_shape(LSOA) +
  tm_polygons("Age00to04", title = "Aged 0 to 4", palette = "Blues", style = "jenks") +
  tm_layout(legend.title.size = 0.8)
```





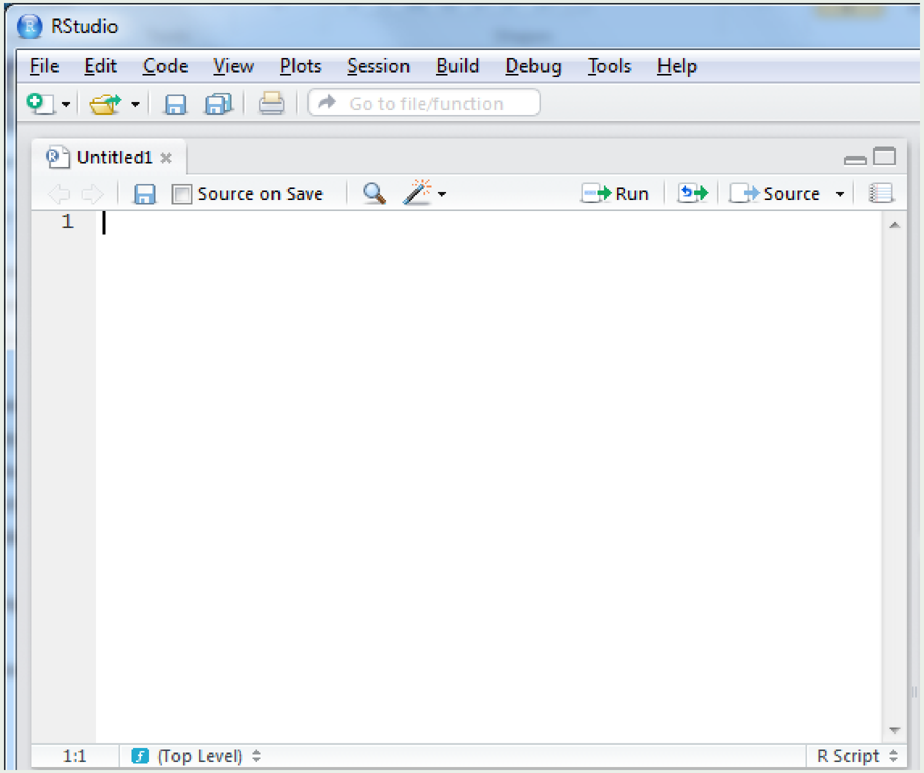
```
tm_shape(LSOA) +
  tm_polygons("Age00to04", title = "Aged 0 to 4", palette = "Blues", style = "jenks") +
  tm_layout(legend.title.size = 0.8)
```

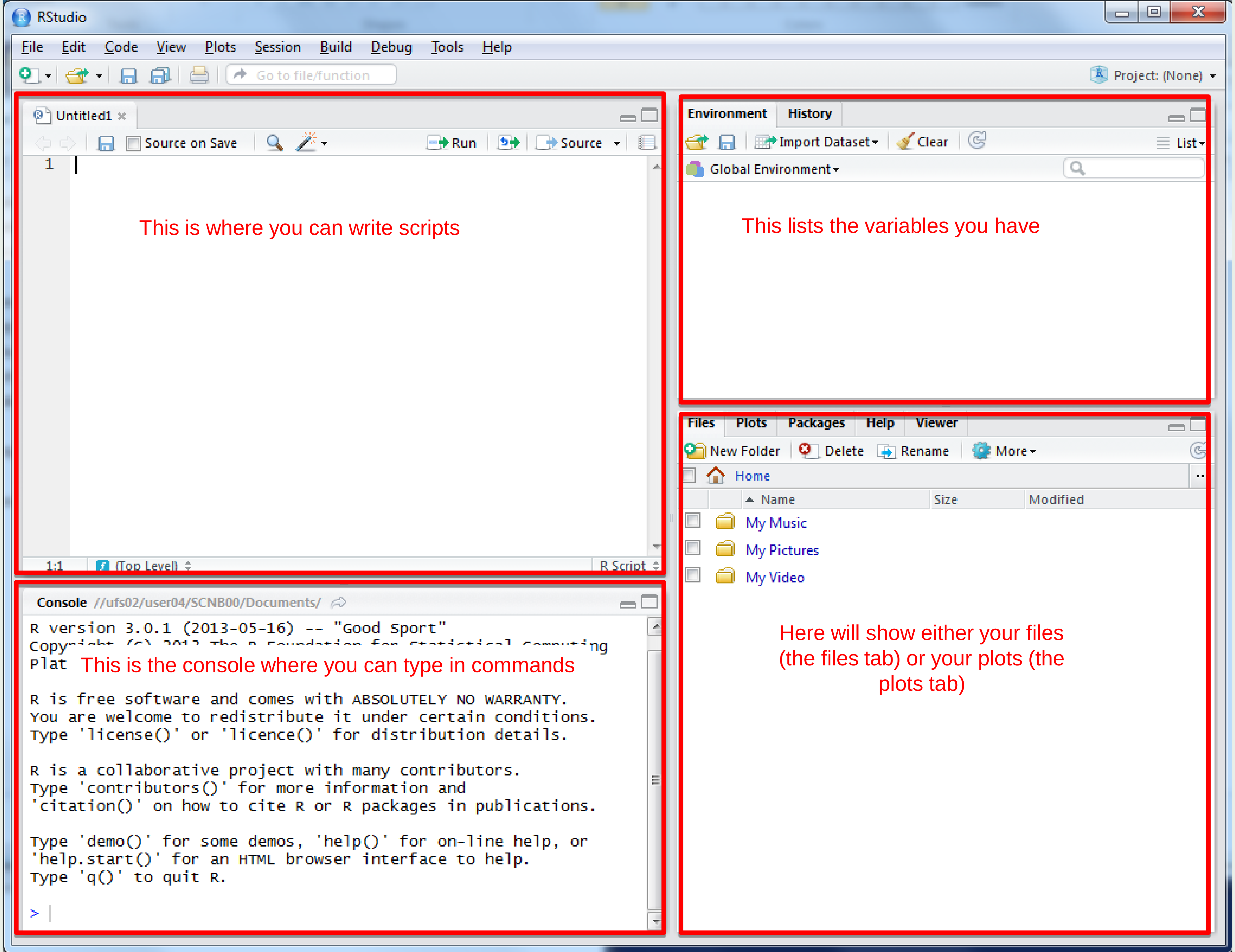


chosen style: one of "fixed", "sd", "equal", "pretty", "quantile", "kmeans", "hclust", "bclust", "fisher", or "jenks"

R Scripts

- Why use scripts?
  - Easier to correct code
  - Can easily re run sections of code, or all code if you need to start again
  - Easy to share code





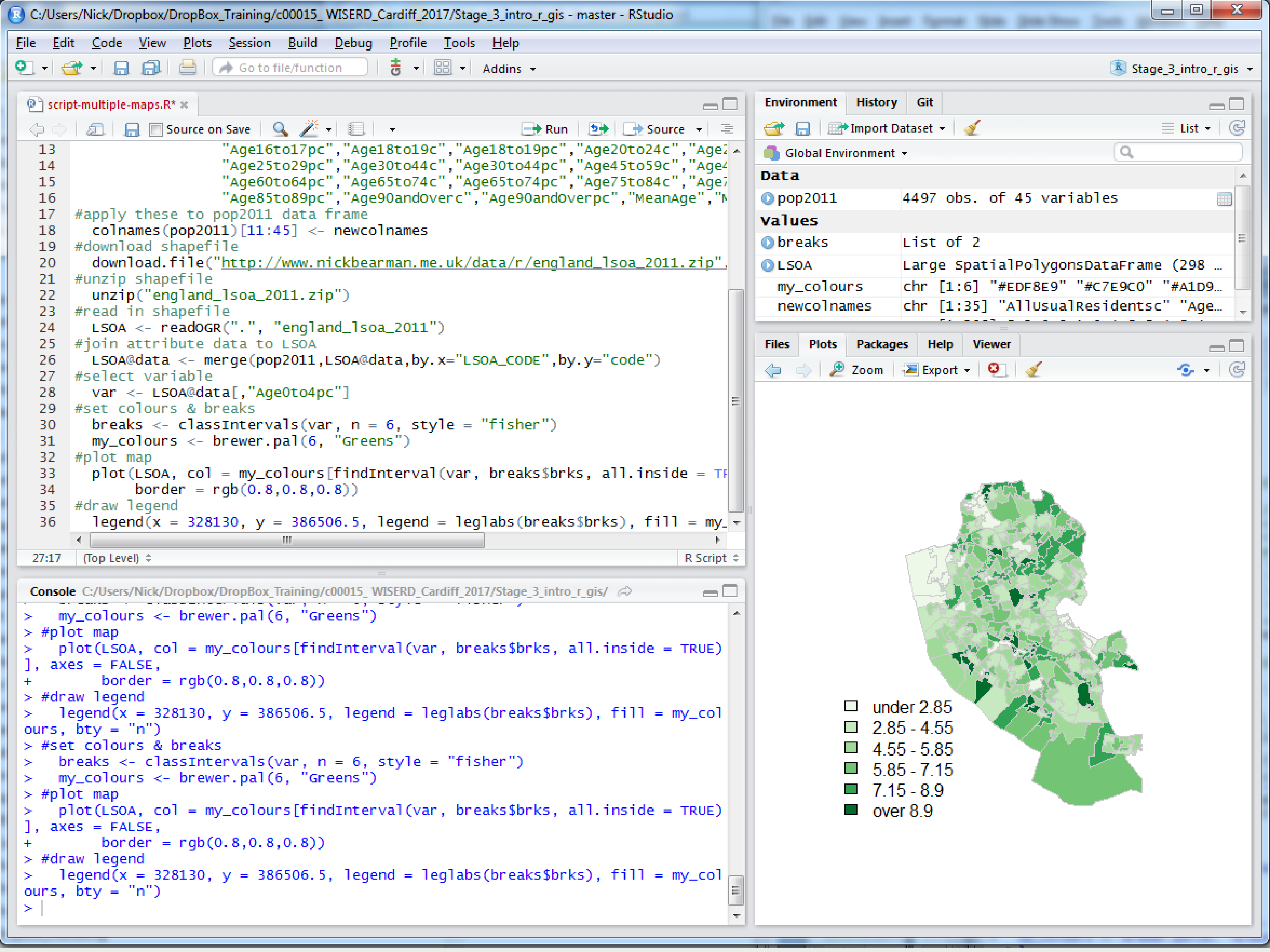
This is where you can write scripts

This lists the variables you have

This is the console where you can type in commands

Here will show either your files (the files tab) or your plots (the plots tab)

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Plat

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```


C:/Users/Nick/Dropbox/DropBox_Training/c00015_WISERD_Cardiff_2017/Stage_3_intro_r_gis - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Stage_3_intro_r_gis
script-multiple-maps.R*
Source on Save
Run
Source
27:17
(Top Level)
R Script
Console C:/Users/Nick/Dropbox/DropBox_Training/c00015_WISERD_Cardiff_2017/Stage_3_intro_r_gis/
Environment History Git
Import Dataset
List
Global Environment
Files Plots Packages Help Viewer
Zoom
Export
□ under 2.85
🟩 2.85 - 4.55
🟩 4.55 - 5.85
🟩 5.85 - 7.15
🟩 7.15 - 8.9
🟩 over 8.9

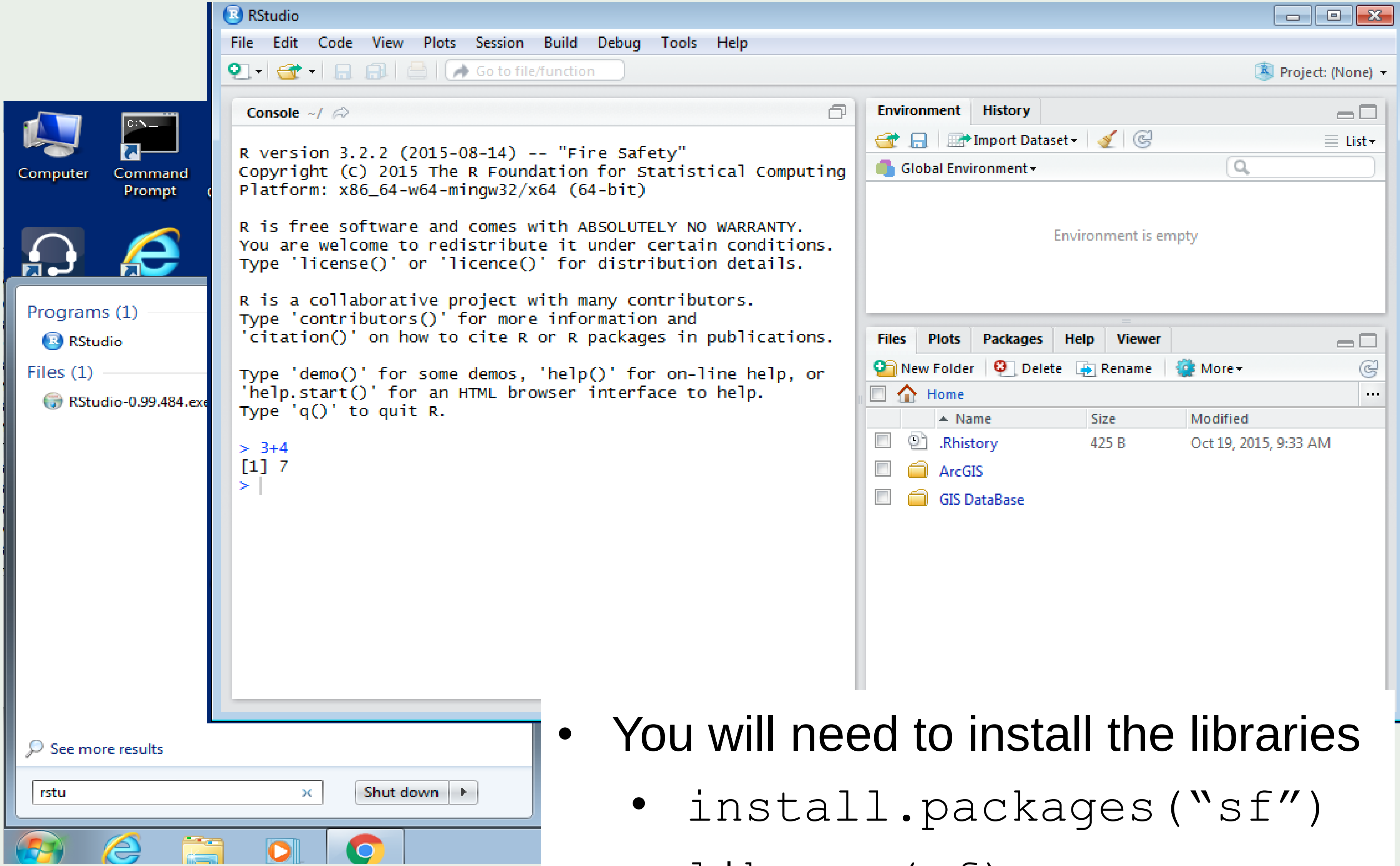
```
                 "Age16to17pc","Age18to19c","Age18to19pc","Age20to24c","Age2
                 "Age25to29pc","Age30to44c","Age30to44pc","Age45to59c","Age4
                 "Age60to64pc","Age65to74c","Age65to74pc","Age75to84c","Age7
                 "Age85to89pc","Age90andOverc","Age90andOverpc","MeanAge","M
#apply these to pop2011 data frame
  colnames(pop2011)[11:45] <- newcolnames
#download shapefile
  download.file("http://www.nickbearman.me.uk/data/r/england_lsoa_2011.zip",
#unzip shapefile
  unzip("england_lsoa_2011.zip")
#read in shapefile
  LSOA <- readOGR(".", "england_lsoa_2011")
#join attribute data to LSOA
  LSOA@data <- merge(pop2011,LSOA@data,by.x="LSOA_CODE",by.y="code")
#select variable
  var <- LSOA@data[,"Age0to4pc"]
#set colours & breaks
  breaks <- classIntervals(var, n = 6, style = "fisher")
  my_colours <- brewer.pal(6, "Greens")
#plot map
  plot(LSOA, col = my_colours[findInterval(var, breaks$brks, all.inside = TR
       border = rgb(0.8,0.8,0.8))
#draw legend
  legend(x = 328130, y = 386506.5, legend = leglabs(breaks$brks), fill = my_
```

```
Data
pop2011        4497 obs. of 45 variables
Values
breaks         List of 2
LSOA           Large SpatialPolygonsDataFrame (298 ...
my_colours     chr [1:6] "#EDF8E9" "#C7E9C0" "#A1D9...
newcolnames    chr [1:35] "AllUsualResidentsc" "Age...
```

```
>   my_colours <- brewer.pal(6, "Greens")
> #plot map
>   plot(LSOA, col = my_colours[findInterval(var, breaks$brks, all.inside = TRUE)
], axes = FALSE,
+        border = rgb(0.8,0.8,0.8))
> #draw legend
>   legend(x = 328130, y = 386506.5, legend = leglabs(breaks$brks), fill = my_col
ours, bty = "n")
> #set colours & breaks
>   breaks <- classIntervals(var, n = 6, style = "fisher")
>   my_colours <- brewer.pal(6, "Greens")
> #plot map
>   plot(LSOA, col = my_colours[findInterval(var, breaks$brks, all.inside = TRUE)
], axes = FALSE,
+        border = rgb(0.8,0.8,0.8))
> #draw legend
>   legend(x = 328130, y = 386506.5, legend = leglabs(breaks$brks), fill = my_col
ours, bty = "n")
> 
```

Load R

Bit.ly/GISRUK-2019

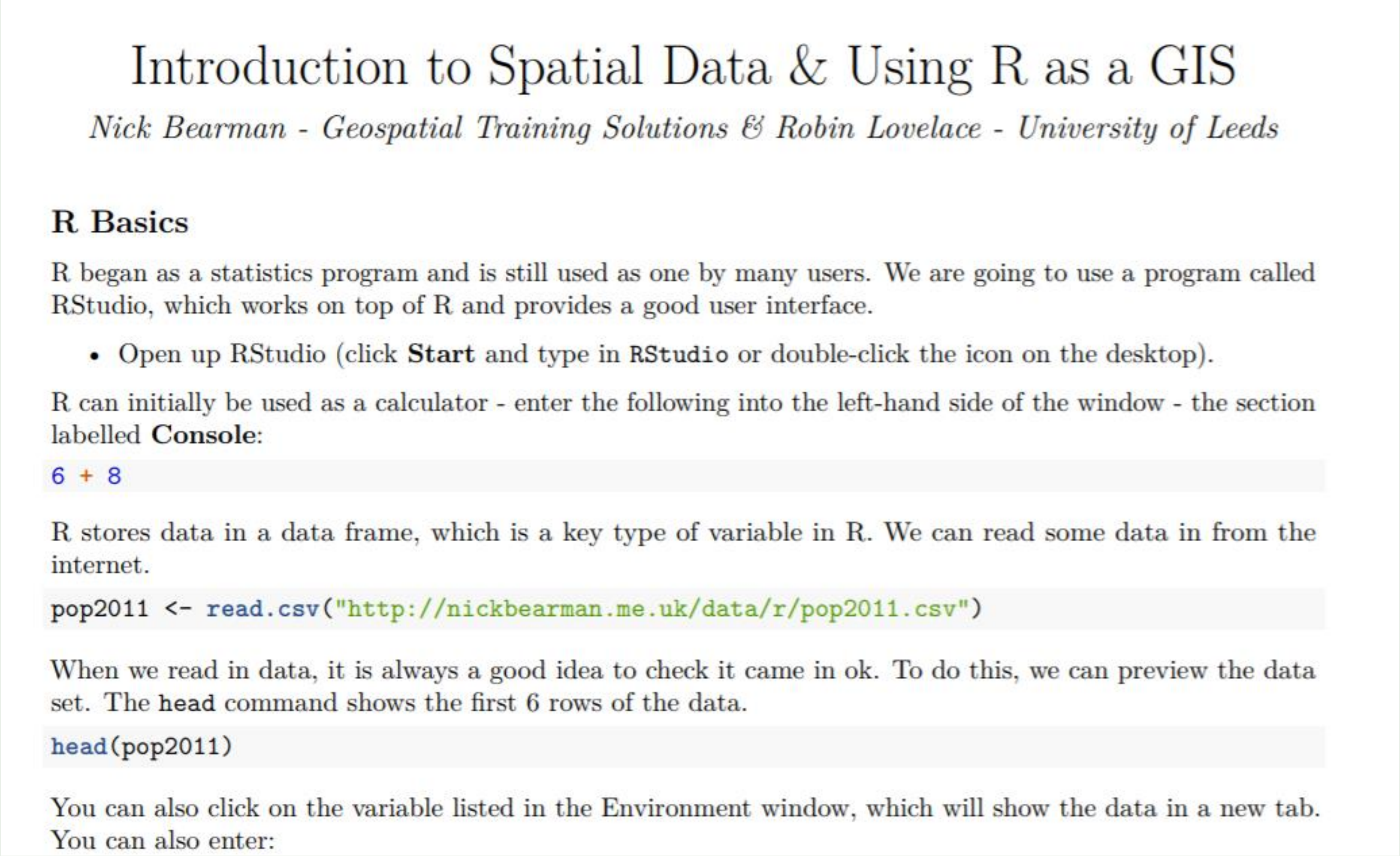


```
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 3+4
[1] 7
> 
```

Name	Size	Modified
.Rhistory	425 B	Oct 19, 2015, 9:33 AM
ArcGIS		
GIS DataBase		

- You will need to install the libraries
    - `install.packages(“sf”)`
    - `library(sf)`

Bit.ly/GISRUK-2019 Workbook.pdf



Introduction to Spatial Data & Using R as a GIS

Nick Bearman - Geospatial Training Solutions & Robin Lovelace - University of Leeds

R Basics

R began as a statistics program and is still used as one by many users. We are going to use a program called RStudio, which works on top of R and provides a good user interface.

- Open up RStudio (click **Start** and type in `RStudio` or double-click the icon on the desktop).

R can initially be used as a calculator - enter the following into the left-hand side of the window - the section labelled **Console**:

```
6 + 8
```

R stores data in a data frame, which is a key type of variable in R. We can read some data in from the internet.

```
pop2011 <- read.csv("http://nickbearman.me.uk/data/r/pop2011.csv")
```

When we read in data, it is always a good idea to check it came in ok. To do this, we can preview the data set. The `head` command shows the first 6 rows of the data.

```
head(pop2011)
```

You can also click on the variable listed in the Environment window, which will show the data in a new tab. You can also enter: