

Rrrrrrr...

An intro to the magic of R!

Getting Set up!

Hopefully everyone will be set up and ready to go!

- R (www.r-project.org)
- R-studio (<https://rstudio.com/products/rstudio/download/>)



WARNING.....

this presentation will contain coding



Quiz : How much do I know about R?

Rrrrrrrrrr

- What is R?
- What can you do with r?

.....Getting your hands dirty!



What is R?

- A language and program
- Origins in research and data analysis
- Open source
- Flexible
- State of the art+13909 packages
- Extensive support & community

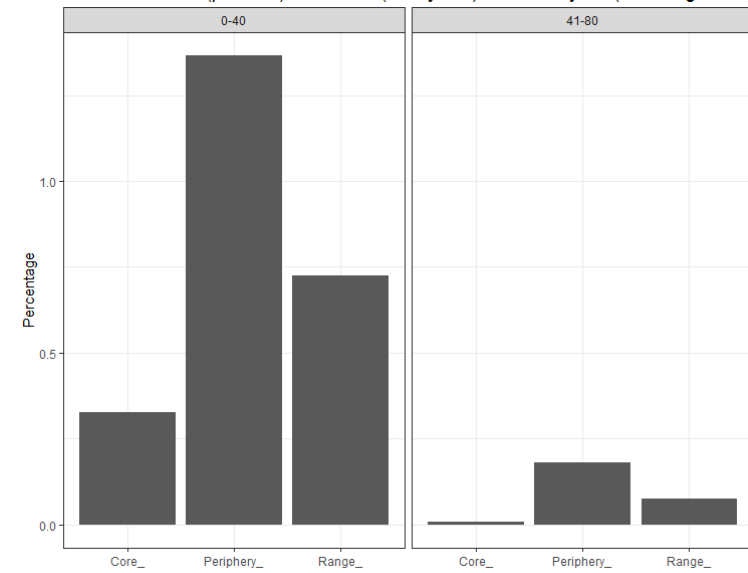


What can you do with R?

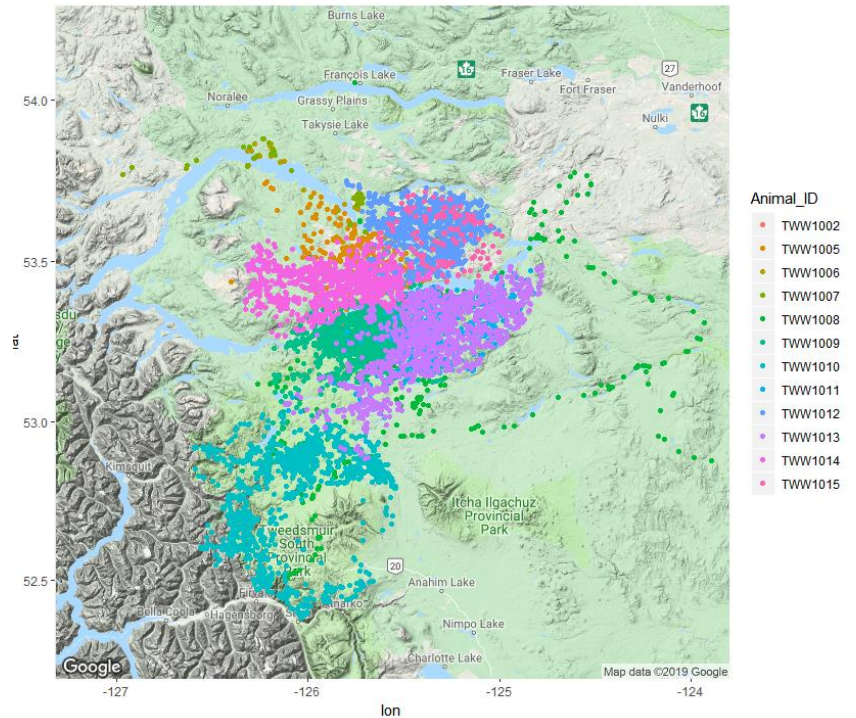
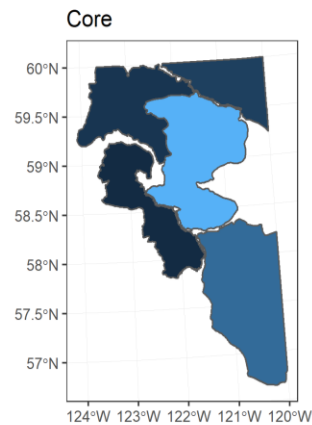
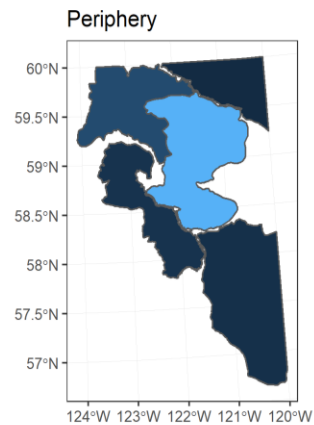
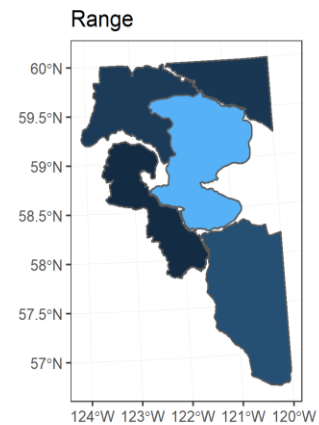
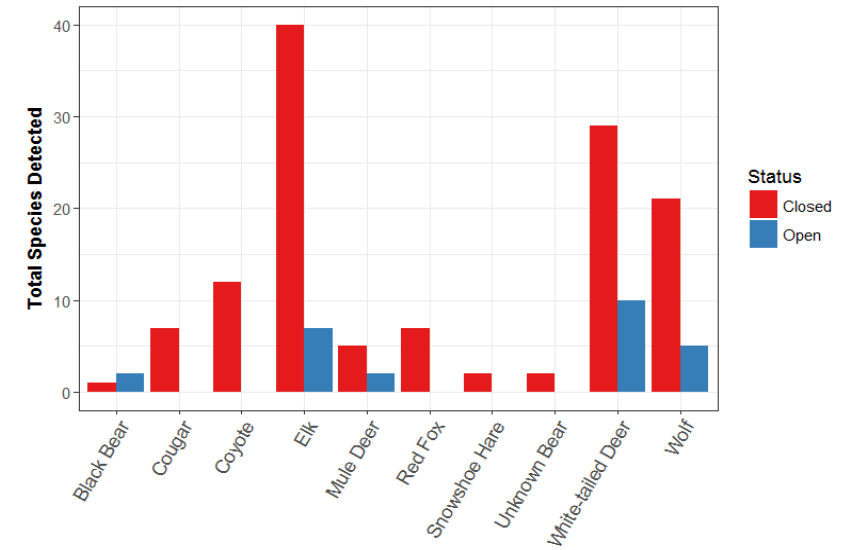
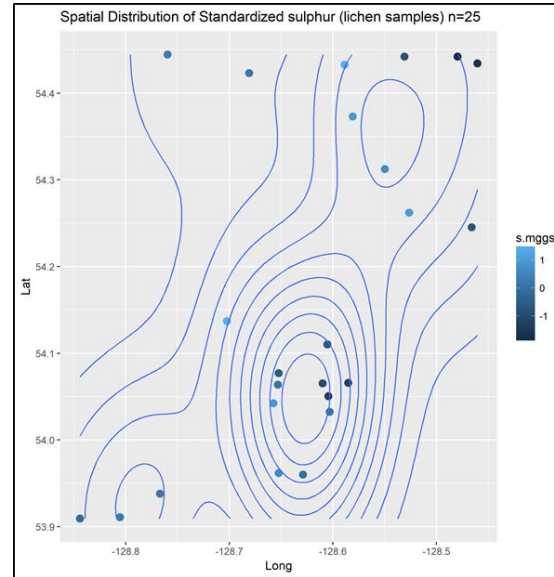
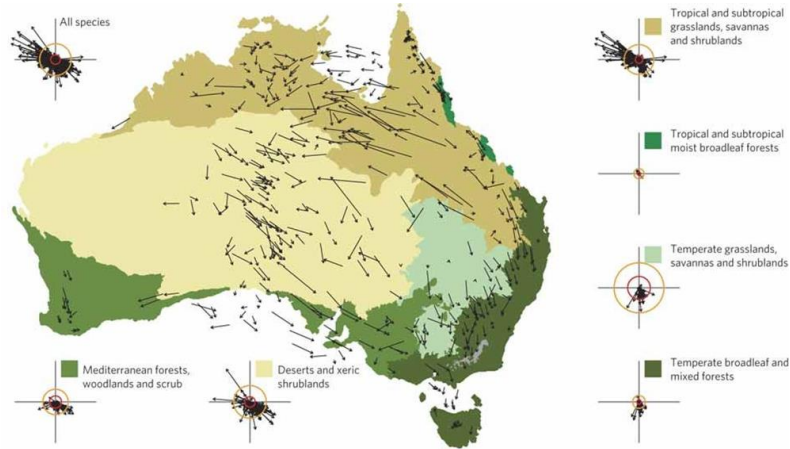
Data manipulation and error checking

The image displays a collage of several overlapping Excel spreadsheets. The primary spreadsheet in the foreground shows a data table with the following columns: ID, lat, lon, year, obs, elev, and sun. The data is organized into rows, with some rows highlighted in green. The spreadsheets in the background show similar data structures, suggesting a process of data manipulation or error checking across multiple datasets.

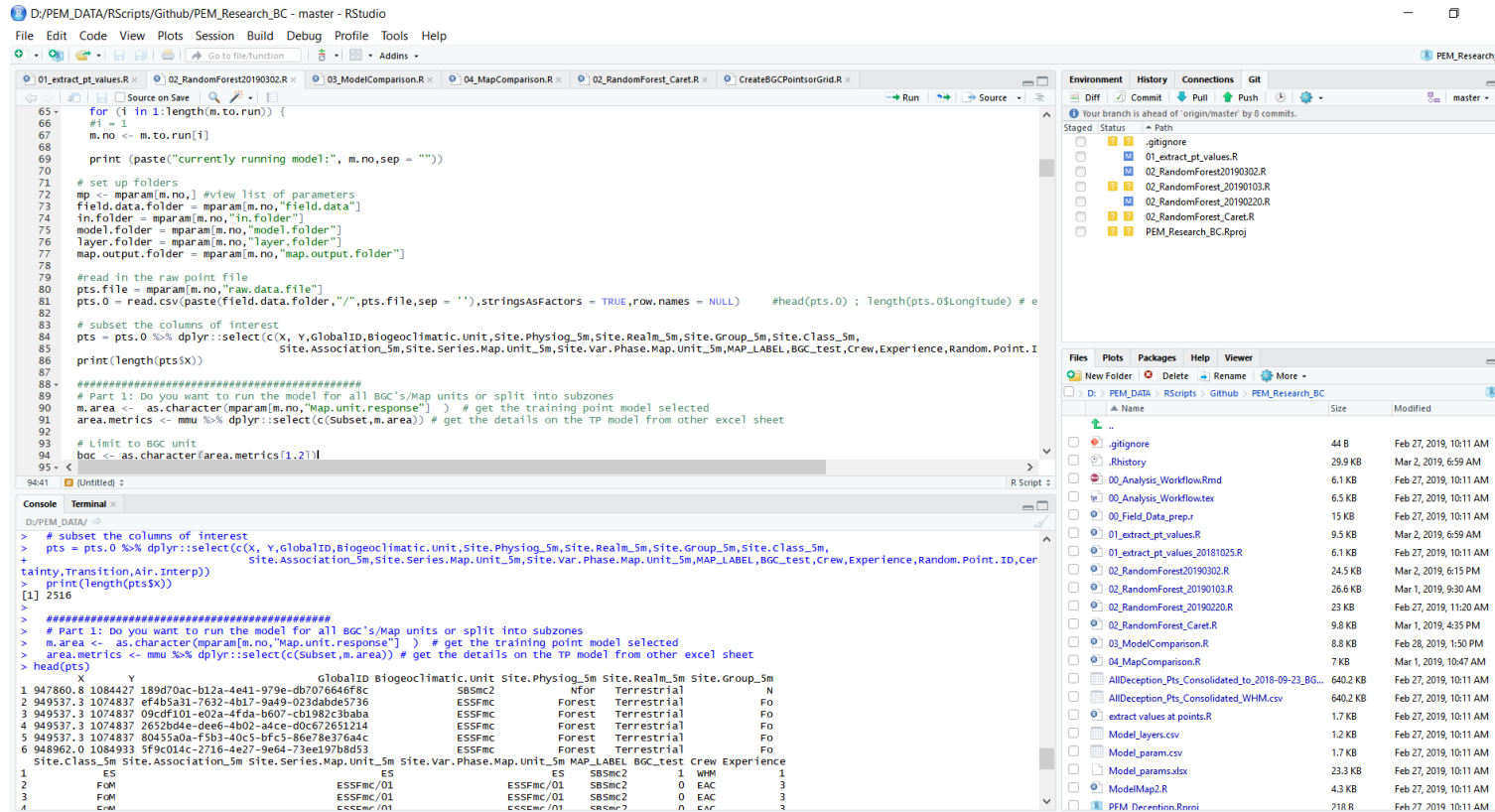
ID	lat	lon	year	obs	elev	sun	
1	17	331	2002	100	1662	56.64286	
2	17	331	2002	83	1707	53.11704	
3	17	331	2002	86	1846	50.67954	
4	17	331	2002	142	1553	53.36312	
5	17	331	2002	126	1876	57.22138	
6	17	331	2002	181	1076	50.65789	
7	17	331	2002	171	1030	57.27132	
8	17	331	2002	160	904	58.87759	
9	17	331	2002	134	1345	58.87459	
10	17	331	2002	86	1110	50.78275	
11	17	331	2002	136	1412	58.30864	
12	17	331	2002	148	960	54.68672	
13	17	331	2002	148	1110	50.78275	
14	17	331	2002	136	1412	58.30864	
15	17	331	2002	93	705	55.75635	
16	17	331	2002	114	1055	54.55151	
17	17	331	2002	137	1008	50.50823	
18	17	331	2002	148	1670	50.55384	
19	17	331	2002	122	822	50.69557	
20	17	331	2002	97	1610	57.60428	
21	17	331	2002	103	1635	55.30886	
22	17	331	2002	88	342	57.27132	
23	17	331	2002	120	759	59.69845	
24	17	331	2002	101	565	55.63962	
25	17	331	2002	142	769	58.97775	
26	17	331	2002	81	1779	802	50.29996
27	17	331	2002	112	1048	51.44035	
28	17	331	2002	105	405	56.47638	
29	17	331	2002	126	1933	51.67409	



Graphing and Visualisation



Analysis : Statistical



The screenshot displays the RStudio interface. The main editor shows R code for a workflow involving data extraction, model fitting, and file management. The console at the bottom shows the execution of the code, including the output of a data frame with 6 columns: X, Y, GlobalID, Biogeoclimatic.Unit, Site, and PhysioLog. The file explorer on the right shows a directory structure for 'PEM_RESEARCH_BC' with various subfolders and files, including '01_extract_pt_values.R', '02_RandomForest_20190302.R', '03_ModelComparison.R', '04_MapComparison.R', '02_RandomForest_Caret.R', and 'CreateBGCPoinGrid.R'.

```
65 for (i in 1:length(m.to.run)) {
66   #i = 1
67   m.no <- m.to.run[i]
68
69   print(paste("currently running model:", m.no, sep = ""))
70
71   # set up folders
72   mp <- mparam[m.no,] #view list of parameters
73   field.data.folder = mparam[m.no,"field.data"]
74   in.folder = mparam[m.no,"in.folder"]
75   model.folder = mparam[m.no,"model.folder"]
76   layer.folder = mparam[m.no,"layer.folder"]
77   map.output.folder = mparam[m.no,"map.output.folder"]
78
79   #read in the raw point file
80   pts.file = mparam[m.no,"raw.data.file"]
81   pts.0 = read.csv(paste(field.data.folder,"/",pts.file,sep = ""),stringsAsFactors = TRUE,row.names = NULL) #head(pts.0) ; length(pts.0$Longitude) # e
82
83   # subset the columns of interest
84   pts = pts.0 %>% dplyr::select(c(X, Y,GlobalID,Biogeoclimatic.Unit,Site,PhysioLog,Site.RealM,Site.Group,Site.Class,Site,
85     Site.Association,Site.Series,Map.Unit,Site.Var.Phase,Map.Unit,MAP_LABEL,BGC_test,Crew,Experience,Random.Point.ID,Cer
86   print(length(pts$X))
87
88   #####
89   # Part 1: do you want to run the model for all BGC's/Map units or split into subzones
90   m.area <- as.character(mparam[m.no,"Map.unit.response"]) # get the training point model selected
91   area.metrics <- mmu %>% dplyr::select(c(Subset,m.area)) # get the details on the TP model from other excel sheet
92
93   # Limit to BGC unit
94   boc <- as.character(area.metrics[1,2])
95 }
```

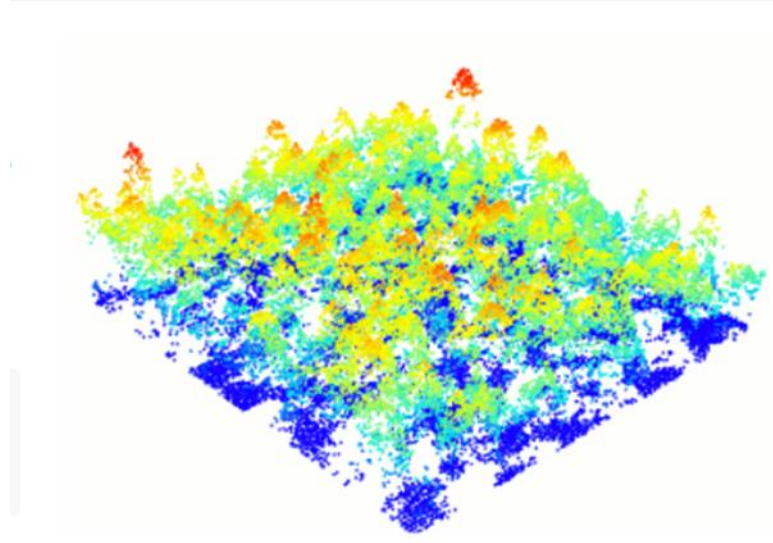
Console Output:

```
> # subset the columns of interest
> pts = pts.0 %>% dplyr::select(c(X, Y,GlobalID,Biogeoclimatic.Unit,Site,PhysioLog,Site.RealM,Site.Group,Site.Class,Site,
+ Site.Association,Site.Series,Map.Unit,Site.Var.Phase,Map.Unit,MAP_LABEL,BGC_test,Crew,Experience,Random.Point.ID,Cer
+ tainty,Transition,Air,Interp))
> print(length(pts$X))
[1] 2516
>
> #####
> # Part 1: do you want to run the model for all BGC's/Map units or split into subzones
> m.area <- as.character(mparam[m.no,"Map.unit.response"]) # get the training point model selected
> area.metrics <- mmu %>% dplyr::select(c(Subset,m.area)) # get the details on the TP model from other excel sheet
> head(pts)
  X Y GlobalID Biogeoclimatic.Unit Site.PhysioLog Site.RealM Site.Group Site.Class
1 947860.8 1084427 189d70ac-b12a-4e41-979e-d0707664f8c SBsmc2 Nfor Terrestrial N
2 949537.3 1074837 ef4b5a31-7632-4b17-9a49-023dabde5736 ESSFmc Forest Terrestrial Fo
3 949537.3 1074837 09cdf101-e02a-4fda-b607-c01982c3ababa ESSFmc Forest Terrestrial Fo
4 949537.3 1074837 2652bda4-dee6-4b02-a4ce-d0c672651214 ESSFmc Forest Terrestrial Fo
5 949537.3 1074837 80455a0a-f5b3-40c5-bfcs-86e78e376a4c ESSFmc Forest Terrestrial Fo
6 948962.0 1084933 5f9c014c-2716-4e27-9e64-73ee197b8d53 ESSFmc Forest Terrestrial Fo
  Site.Class Site.Association Site.Series Map.Unit Site.Var.Phase Map.Unit MAP_LABEL BGC_test Crew Experience
1 ES ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01
2 ES ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01
3 ES ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01
4 ES ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01 ESSFmc/01
```

- Regression
- Mixed models
- Random forest
- Bayesian statistical
- Time-lapse Analysis
- Classification/ Machine Learning

Spatial Analysis /GIS

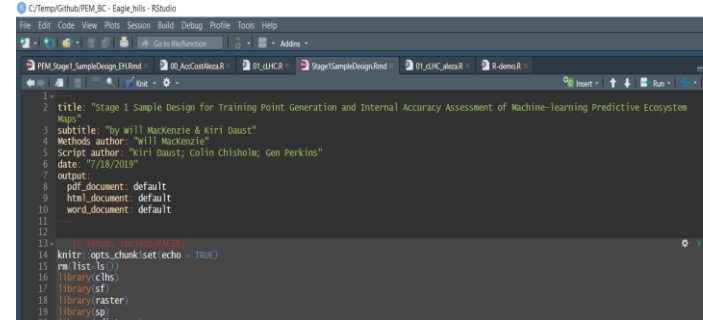
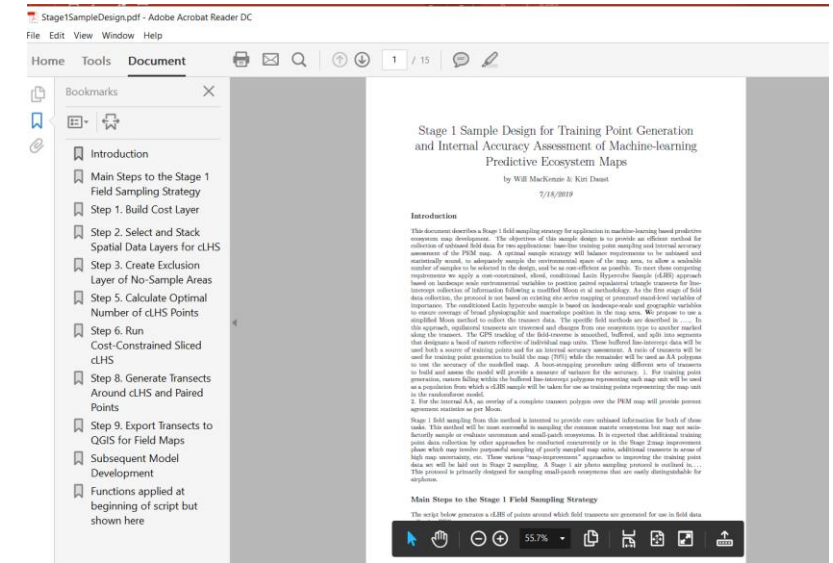
- Shapefiles / Geopackage
- Maps
- KML
- Lidar data
- Sampling Plans



..... Stay tuned for the Demo!

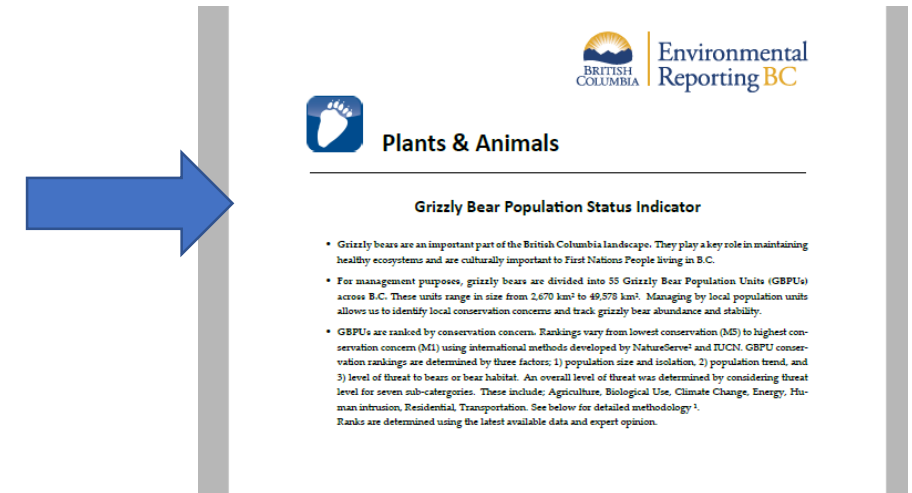
Outputs: Documents

- R markdown to script your outputs
- Word
- Powerpoints / Slides



Grizzly Bear Population Status Indicator

- Grizzly bears are an important part of the British Columbia landscape. They play a key role in maintaining healthy ecosystems and are culturally important to First Nations People living in B.C.
- For management purposes, grizzly bears are divided into 55 Grizzly Bear Population Units (GBPUs) across B.C. These units range in size from 2,670 km² to 49,578 km². Managing by local population units allows us to identify local conservation concerns and track grizzly bear abundance and stability.



Outputs: Interactive data display

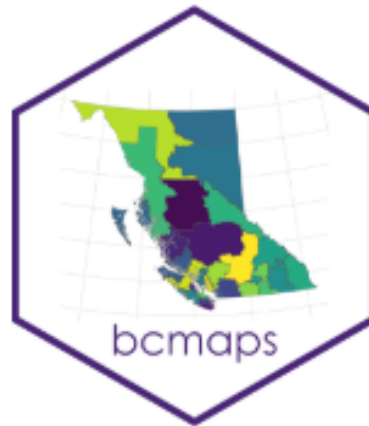
- Shiny App / leaflet mapping
 - Roads and Roadless Areas in British Columbia

<http://www.env.gov.bc.ca/soe/indicators/land/roads.html>

Packages to use provincial datasets

BC gov specific packages :

- bcmaps
- bcdata
- tidyhydat



Why Do I Use R ?

The Bad

Steep learning curve

... but lots of resources available

Can be slow with large datasets

... but work arounds (postgres/packages)

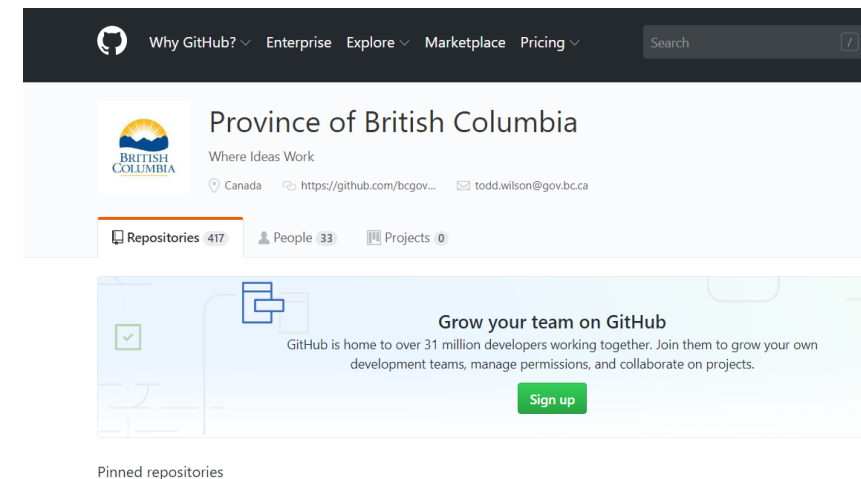


Why Do I Use R ?



The Good

- Efficiency !!
- Transparency & documentation
- Repeatability
- Flexible and ever expanding
- Community to share code – Github BCGov
- Compatible with other software



Lets get started.....

