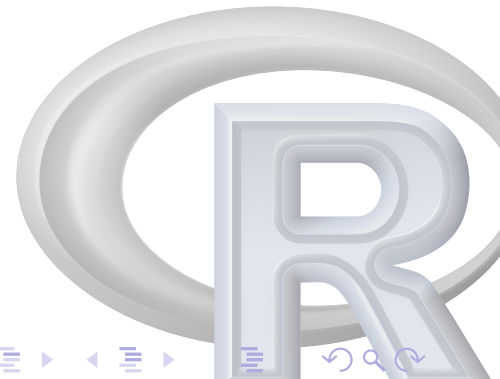
Seminar One:
An Introduction to R

September 9, 2014



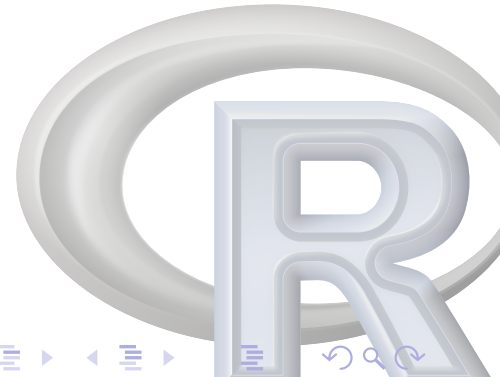
Seminars

- **Seminar 1:** Introduction to R
- **Seminar 2:** Data manipulation
- **Seminar 3:** Functions
- **Seminar 4:** Basic statistical models
- **Seminar 5:** Plotting in R—base package
- **Seminar 6:** Plotting in R—ggplot

- LU: Hannah Franklin, Maki Ikegami, Jennifer Bufford, Ursula Torres, Marona Capdevila, Samuel Brown
- UC: Olivia Burge, Rachel Harley, Simon Howard, Jon Bray, Laura Azzani, Camille Coux, Helen Warburton, Richard White

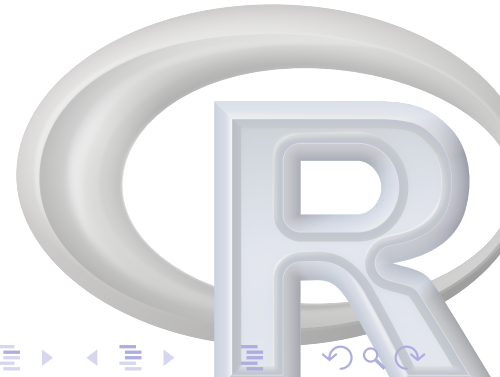
What is R?

- Environment for data analysis
- Free
- Powerful
- Flexible
- Responsive
- Ubiquitous
- Handle all types of data



Getting started

- Download from www.r-project.org
- Don't panic!
- `5 + 3`
- `23 * 56`
- `aa <- c(1, 3, 6, 9)`
- `mean(aa)`
- `aa + 3`



Terminology 1

Object: Anything stored in R memory

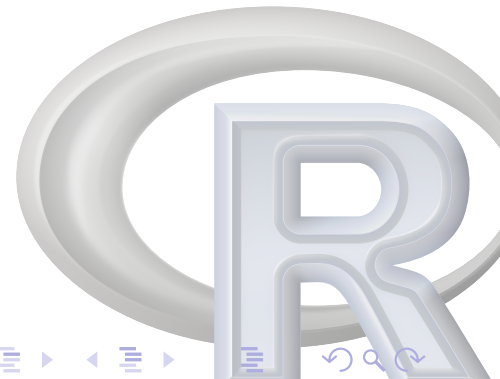
Function: An object that does something; a process; characterised by parentheses

Vector: A one-dimensional object with all elements of a single mode

e.g.
```
1 2 4 8 16 32
"this" "is" "a" "vector"
```

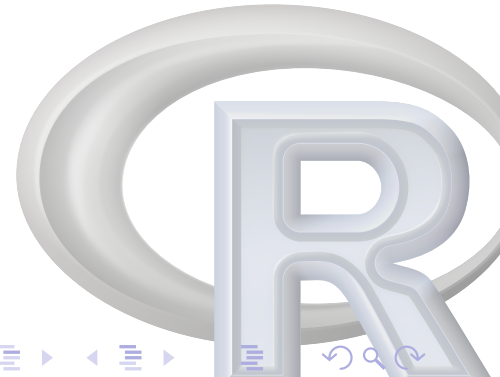
Modes: The form of the object

- **Numeric:** Numbers
- **Character:** Strings
- **Factor:** Categories
- **Logical:** TRUE and FALSE



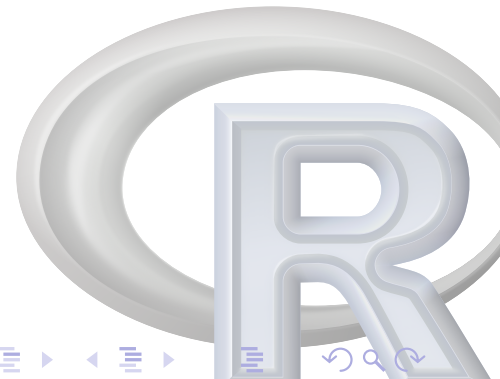
RStudio

- Integrated Development Environment (organisation system)
- Scripts
- Projects
- Help files
- Package loading



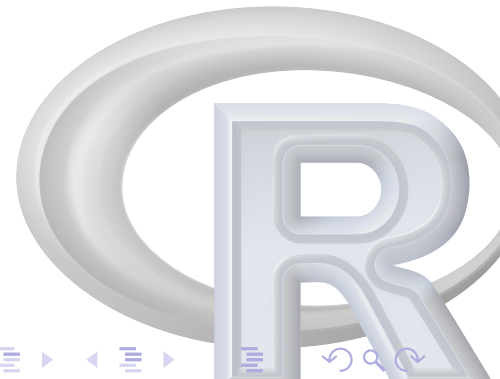
Getting data in

```
getwd()
setwd()
read.table()
read.csv()
file.choose()
```



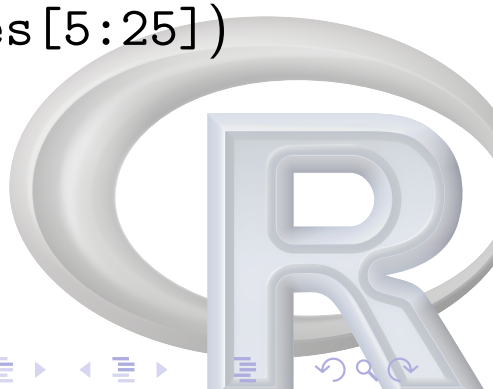
Terminology 2

Matrix:	A two-dimensional object with all elements of a single mode
List:	A collection of objects
Data frame:	A list with all objects being vectors of the same length; a spreadsheet
#:	Comment character. Everything after # is ignored by R
Indexing:	Locating and extracting values within an object



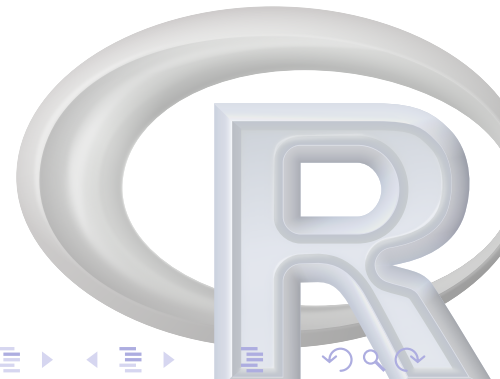
Indexing

- Square bracket notation:
  - Single value in vectors and matrices (`aa[3]`)
  - Two values (row, column) in matrices and dataframes (`iris[3, 4]`)
  - Ranges can be given (`iris[1:20, 4]`)
  - A vector object can be given (`bb <- c(3, 5, 8:9, 13); iris[bb, 2:3]`)
  - Whole rows and columns can be extracted, but comma must be present (`iris[, 5]`)
- Dollar notation:
  - Used in lists and dataframes
  - The name of the variable is used (`iris$Species`)
  - Can be combined with the above (`iris$Species[5:25]`)
- @ notation:
  - Used in advanced R objects (S4 classes)
  - Not commonly encountered



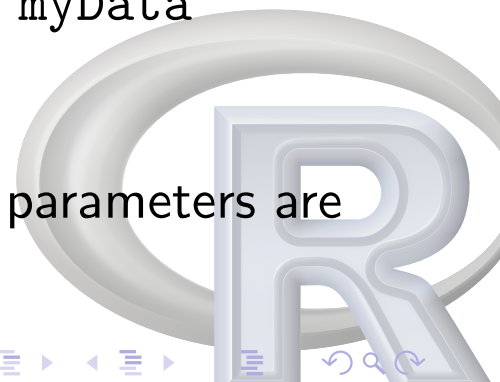
Getting help

- In-program:
  - Specific functions (`?read.table`)
  - Keyword searching (`??`)
  - `example(function)`
  - Package vignettes (`vignette()`)
- Websites:
  - www.r-project.org
  - www.r-bloggers.com
  - www.rseek.org
  - R help email list (search archives first!)
- Books:
  - The R Book by Matt Crawley
  - Use R! series
- People:
  - Seminar organisers
  - Spatial Ecology Group



Other useful things to know

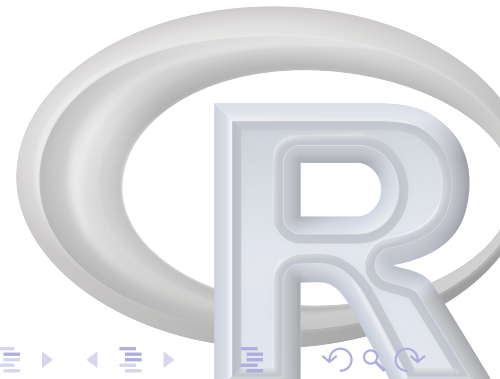
- Packages:
  - User-contributed function sets
  - Dramatically extend base R utility
  - Are the reason for R's popularity
  - Use with caution
  - CRAN (http://cran.r-project.org)
- Missing data:
  - Coded as NA
  - If vector has NA, some functions (eg. `mean()`) will return NA
  - Need to know how to deal with this
  - `mean(xx, na.rm = TRUE)` or `is.na()`
- Syntax:
  - Capitalisation matters! `mydata` is different from `myData`
  - Avoid overnaming functions
  - Whitespace does not matter (usually)
  - Ensure brackets and quotes are closed, function parameters are seperated by commas.



Useful functions

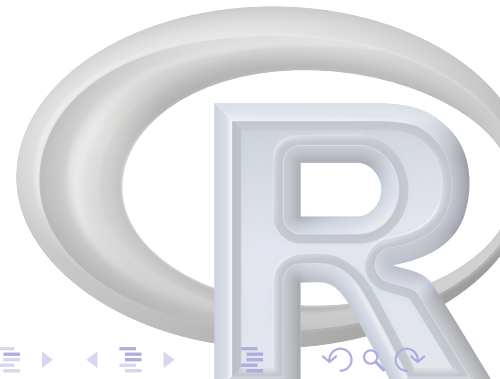
```
summary()
str()
cbind()
rbind()
as.character()
as.factor()
seq()
sample()
rnorm()
runif()
is.na()
match()
which()
```

```
max()
min()
range()
sum()
mean()
median()
cor()
log()
scale()
grep()
gsub()
paste()
strsplit()
```



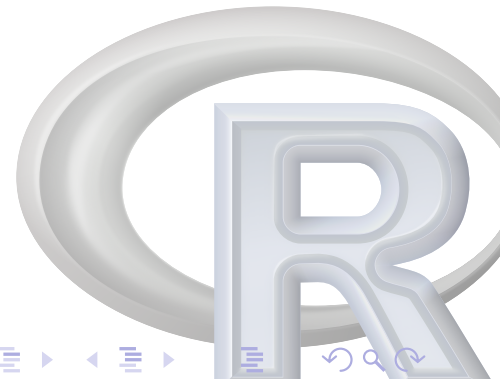
Errors

- `+`
  - Line has not been completed— often missing a close parentheses or quotation mark
- `object 'xxx' not found`
  - Object name has not been assigned
- `could not find function "xxx"`
  - Function does not exist—check spelling, packages loaded, parentheses instead
- `unexpected symbol in "xxxx"`
  - You've probably missed a comma somewhere
- `invalid factor level, NA generated`
  - Tried to add an unrecognised category

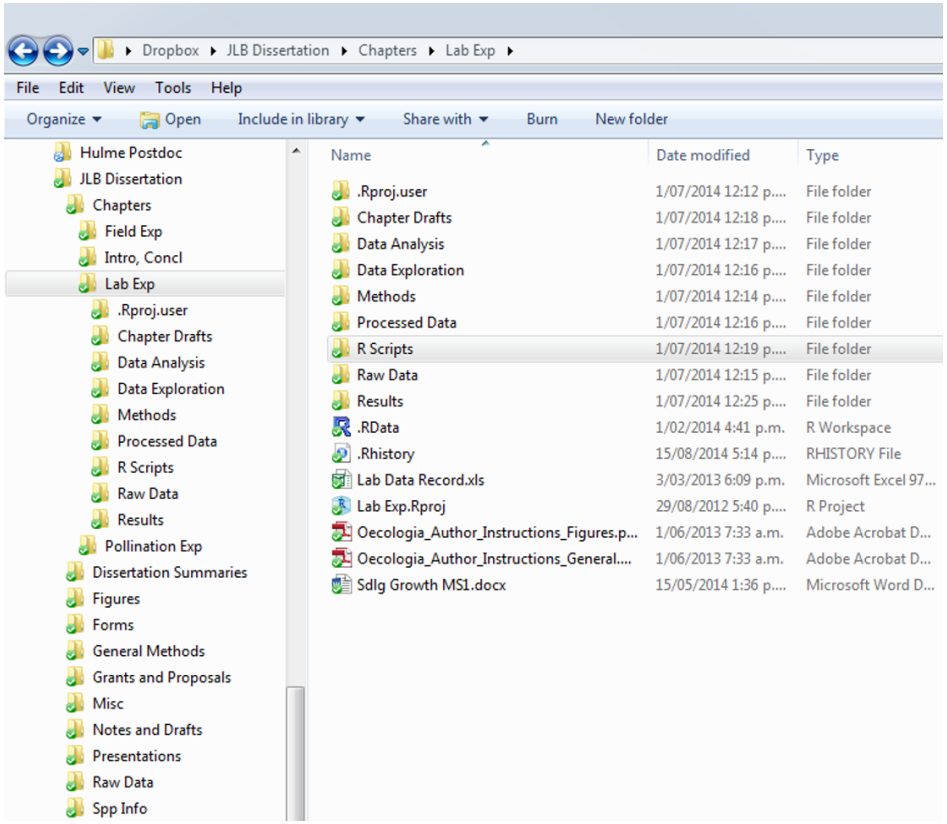


Plotting

```
plot()
hist()
boxplot()
```

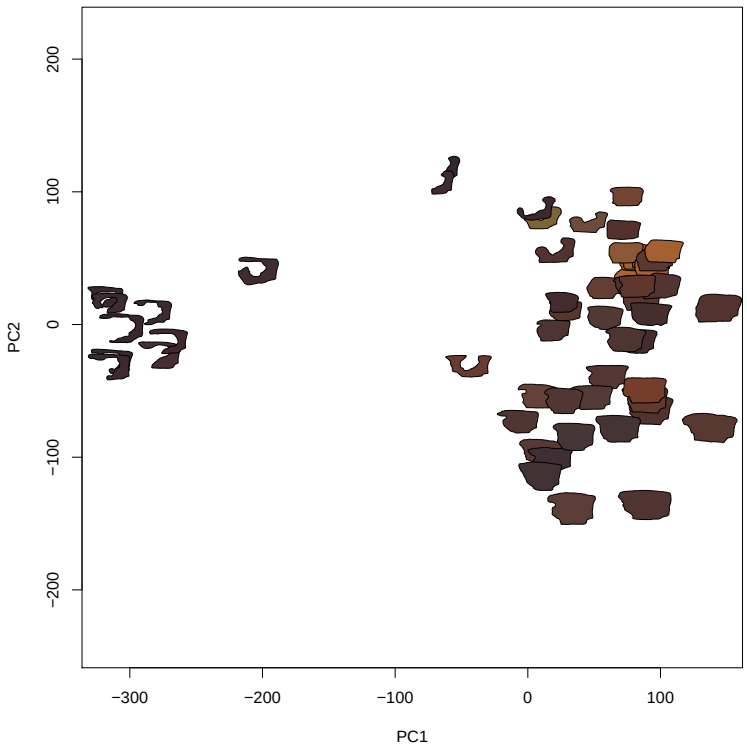


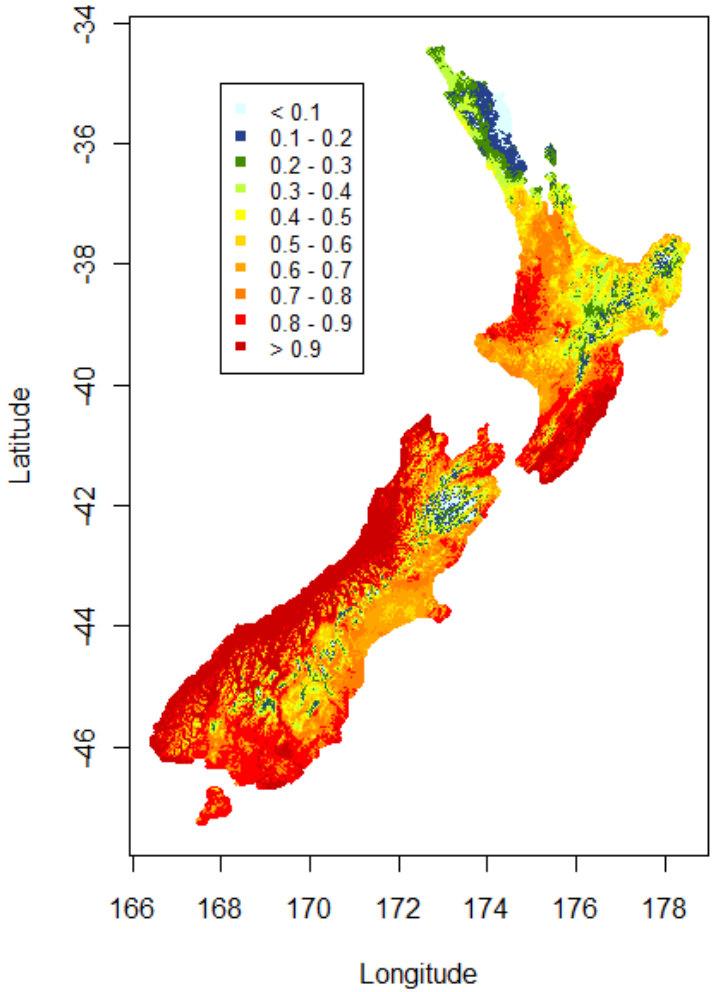
Scripts and organisation



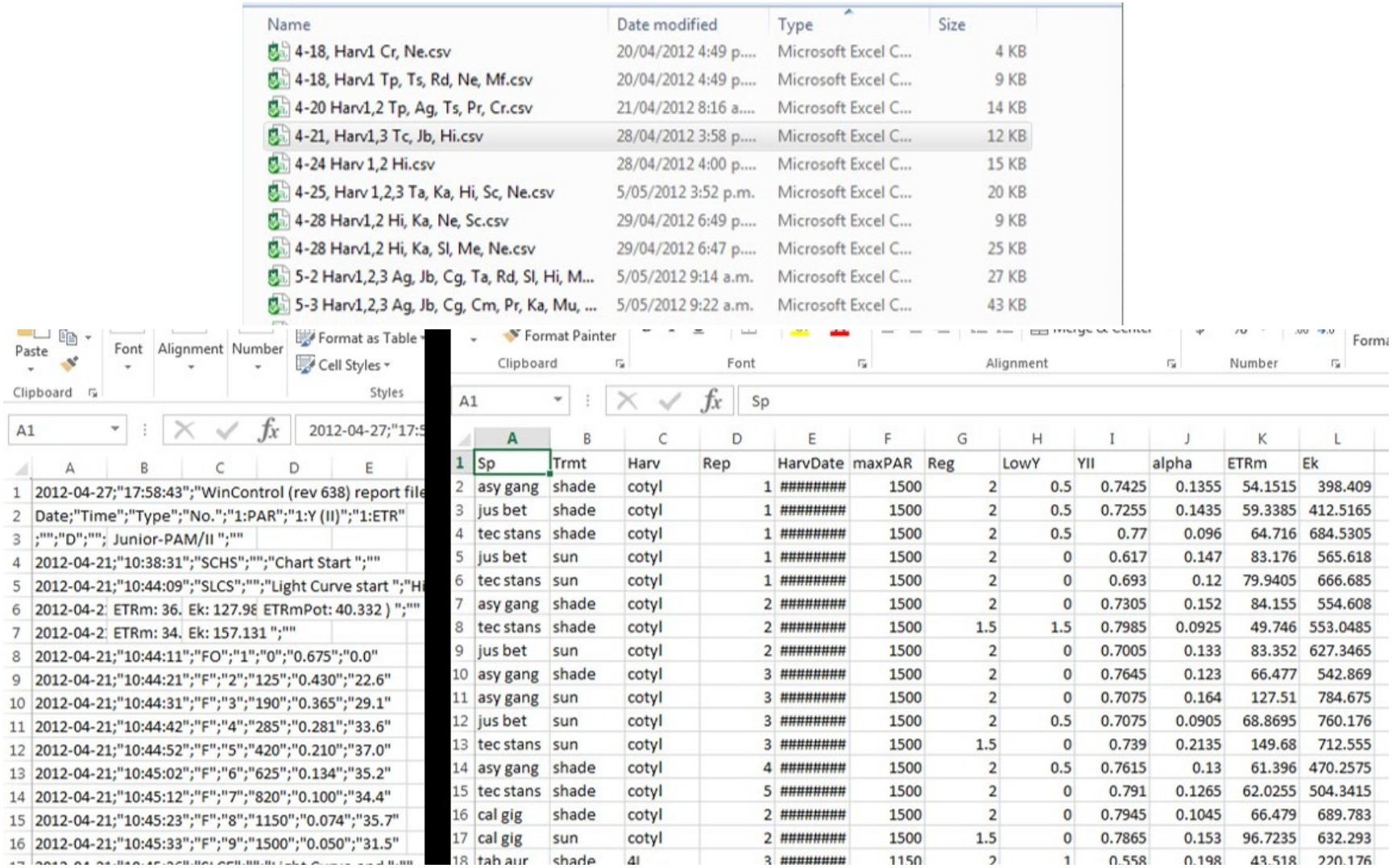

```
######################### Pathogens: Data Exploration ##########################
                            ## Jennifer Bufford ##
                        ## Phil Hulme, Richard Duncan ##
                             # September 2, 2014 #
################################################################################
#Input: Final DB.RData (nzdb)

#Requires: ggplot2, plyr

#Output: Prelim Hosts per Pathogen Graphs.pdf
#        Host Accumulation.pdf
#        Path Data Explor.RData (accum.hosts, nzdbp, FH, FH.yr, Path)
#        Path Data Explor All.RData (all objects)

#Notes
################################################################################

# setwd("C:/Users/Jennifer/Desktop/Dropbox/Hulme Postdoc/Plant-Pathogen DB")

library(ggplot2)
library(plyr)

load('R Scripts/Final DB.RData')

##### Format Data ##############################################################

## Select Data ##

nzdbp <- nzdb[!is.na(nzdb$FSp) & !is.na(nzdb$HSp),]
nzdbp <- nzdbp[nzdbp$FStat %in% "p",] #14103 lines, >17000 records
#excludes Phytophthora, other unmarked 'fungi'

## Format Biostatus ##

nzdbp[grep('Exotic', nzdbp$HBio), 'HBio2'] <- "Exotic"
nzdbp[grepl('Indigenous|Endemic|Non-endemic', nzdbp$HBio) & is.na(nzdbp$HBio2),'HBio2'] <-
  "Native"

nzdbp[grep('Exotic', nzdbp$FBio), 'FBio2'] <- "Exotic"
nzdbp[grepl('Indigenous|Endemic|Non-endemic', nzdbp$FBio) & is.na(nzdbp$FBio2),'FBio2'] <-
```

Examples of R use





Examples of R use



Name	Date modified	Type	Size
4-18, Harv1 Cr, Ne.csv	20/04/2012 4:49 p....	Microsoft Excel C...	4 KB
4-18, Harv1 Tp, Ts, Rd, Ne, Mf.csv	20/04/2012 4:49 p....	Microsoft Excel C...	9 KB
4-20 Harv1,2 Tp, Ag, Ts, Pr, Cr.csv	21/04/2012 8:16 a....	Microsoft Excel C...	14 KB
4-21, Harv1,3 Tc, Jb, Hi.csv	28/04/2012 3:58 p....	Microsoft Excel C...	12 KB
4-24 Harv 1,2 Hi.csv	28/04/2012 4:00 p....	Microsoft Excel C...	15 KB
4-25, Harv 1,2,3 Ta, Ka, Hi, Sc, Ne.csv	5/05/2012 3:52 p.m.	Microsoft Excel C...	20 KB
4-28 Harv1,2 Hi, Ka, Ne, Sc.csv	29/04/2012 6:49 p....	Microsoft Excel C...	9 KB
4-28 Harv1,2 Hi, Ka, Sl, Me, Ne.csv	29/04/2012 6:47 p....	Microsoft Excel C...	25 KB
5-2 Harv1,2,3 Ag, Jb, Cg, Ta, Rd, Sl, Hi, M...	5/05/2012 9:14 a.m.	Microsoft Excel C...	27 KB
5-3 Harv1,2,3 Ag, Jb, Cg, Cm, Pr, Ka, Mu, ...	5/05/2012 9:22 a.m.	Microsoft Excel C...	43 KB

	A	B	C	D	E
1	2012-04-27;"17:58:43";"WinControl (rev 638) report file				
2	Date;"Time";"Type";"No.";"1:PAR";"1:Y (II)";"1:ETR"				
3	;"";"D";""; Junior-PAM/II ";""				
4	2012-04-21;"10:38:31";"SCHS";"";"Chart Start ";""				
5	2012-04-21;"10:44:09";"SLCS";"";"Light Curve start ";"Hi				
6	2012-04-2 ETRm: 36. Ek: 127.98 ETRmPot: 40.332) ";""				
7	2012-04-2 ETRm: 34. Ek: 157.131 ";""				
8	2012-04-21;"10:44:11";"FO";"1";"0";"0.675";"0.0"				
9	2012-04-21;"10:44:21";"F";"2";"125";"0.430";"22.6"				
10	2012-04-21;"10:44:31";"F";"3";"190";"0.365";"29.1"				
11	2012-04-21;"10:44:42";"F";"4";"285";"0.281";"33.6"				
12	2012-04-21;"10:44:52";"F";"5";"420";"0.210";"37.0"				
13	2012-04-21;"10:45:02";"F";"6";"625";"0.134";"35.2"				
14	2012-04-21;"10:45:12";"F";"7";"820";"0.100";"34.4"				
15	2012-04-21;"10:45:23";"F";"8";"1150";"0.074";"35.7"				
16	2012-04-21;"10:45:33";"F";"9";"1500";"0.050";"31.5"				

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sp	Trmt	Harv	Rep	HarvDate	maxPAR	Reg	LowY	YII	alpha	ETRm	Ek
2	asy gang	shade	cotyl	1	########	1500	2	0.5	0.7425	0.1355	54.1515	398.409
3	jus bet	shade	cotyl	1	########	1500	2	0.5	0.7255	0.1435	59.3385	412.5165
4	tec stans	shade	cotyl	1	########	1500	2	0.5	0.77	0.096	64.716	684.5305
5	jus bet	sun	cotyl	1	########	1500	2	0	0.617	0.147	83.176	565.618
6	tec stans	sun	cotyl	1	########	1500	2	0	0.693	0.12	79.9405	666.685
7	asy gang	shade	cotyl	2	########	1500	2	0	0.7305	0.152	84.155	554.608
8	tec stans	shade	cotyl	2	########	1500	1.5	1.5	0.7985	0.0925	49.746	553.0485
9	jus bet	sun	cotyl	2	########	1500	2	0	0.7005	0.133	83.352	627.3465
10	asy gang	shade	cotyl	3	########	1500	2	0	0.7645	0.123	66.477	542.869
11	asy gang	sun	cotyl	3	########	1500	2	0	0.7075	0.164	127.51	784.675
12	jus bet	sun	cotyl	3	########	1500	2	0.5	0.7075	0.0905	68.8695	760.176
13	tec stans	sun	cotyl	3	########	1500	1.5	0	0.739	0.2135	149.68	712.555
14	asy gang	shade	cotyl	4	########	1500	2	0.5	0.7615	0.13	61.396	470.2575
15	tec stans	shade	cotyl	5	########	1500	2	0	0.791	0.1265	62.0255	504.3415
16	cal gig	shade	cotyl	2	########	1500	2	0	0.7945	0.1045	66.479	689.783
17	cal gig	sun	cotyl	2	########	1500	1.5	0	0.7865	0.153	96.7235	632.293
18	tab aur	shade	4l	3	########	1150	2	1	0.558	0.198	43.518	220.176

Thank you!