

Plotting in R

1. Exploratory plots

This week we're going to use a NOAA dataset to explore patterns in reef fish biomass and diversity at 15 Hawaiian islands.

The data were collected by underwater visual census over 2010-2012 - details on the project can be found here:

<http://www.pifsc.noaa.gov/cred/fish>

First thing, as always, is to load in the data frame and explore its structure. Use the commands you've learned in earlier workshops to help you do this.

```
haw<-read.csv(file="/Users/james/Desktop/LearnR/CRED_reef_hawaii_biomass.csv")
str(haw)
dim(haw)
head(haw)
```

In this dataset, we have the biomass (in grams) of different trophic groups (Pisc = piscivores, Herb = herbivores, LowCarn = benthic carnivores, Planktivore = planktivores) at 8 unpopulated islands in the Northwest Hawaiian archipelago (NWHI) and 7 populated islands in the Main Hawaiian Islands (MHI). We also have sea surface temperature (mean annual, SST), productivity (mean annual, chlorophyll a), species richness (total number of species per trophic group) and the log10 human population size (log_population) at each island. The variable 'state' indicates if the island is populated (disturbed) or unpopulated (remote).

Plotting in R is very straightforward - we're going to use the base commands plot(), hist() and boxplot() to explore the reef fish dataset. Before we plot relationships, it's a good idea to explore the distribution of individual variables. Histograms and boxplots are ideal for this.

```
hist(haw$biomass)
hist(haw$biomass, labels = TRUE) ## add labels to bars
hist(haw$richness)
```

The dataset has a lot of categorical variables - we should explore differences across islands and trophic groups.

```
boxplot(haw$biomass ~ haw$trophic) ### Strong differences across trophic groups
boxplot(haw$biomass ~ haw$island)
### Let's examine this relationship for one trophic group using the subset argument
boxplot(haw$biomass ~ haw$island, subset = haw$trophic=="Pisc")
```

Here, we only have one data point per island - the piscivore biomass. We could look at differences across islands based on their human population size (the state and log_population variables).

```
boxplot(haw$biomass ~ haw$state, subset = haw$trophic=="Pisc")
```

There's something going on here - let's delve deeper. What about log_population? We need to examine relationships between continuous variables here...

2. Relationships between continuous variables

Boxplots don't work for relationships between continuous variables. Instead, we want to use plot() - this is the most simple (and powerful) form of plot in R. We ask R to plot X by Y (check with ?plot). Let's examine the relationship between population size and piscivore biomass.

To make our code simpler, we might want to subset our data first..

```
pisc<-haw[haw$trophic=="Pisc",] ### keep only piscivores
```

```
plot(pisc$log_population, pisc$biomass)
plot(pisc$productivity, pisc$biomass)
```

We can still use categorical variables here as well...

```
plot(pisc$state, pisc$biomass)
plot(pisc$trophic, pisc$biomass)
```

So we appear to have a relationship between piscivore biomass and 1) productivity; 2) human population.

We should customise our plots to make them readable (and interesting) for others (like your supervisor).

3. Customizing plots in R

There are many many ways to customise plots in R. ?plot shows just a few examples

```
plot(pisc$productivity, pisc$biomass, col=pisc$state) # Add colour
```

```
plot(pisc$productivity, pisc$biomass, col=pisc$state,
     xlab="Productivity (Chl-a)", ylab="Biomass (grams)", main="Piscivore biomass at Hawaiian islands")
# Add labels to axes
```

```
plot(pisc$productivity, pisc$biomass, col=pisc$state, pch=1,cex=1.5, xlab="Productivity (Chl-a)",
     ylab="Biomass (grams)", main="Piscivore biomass at Hawaiian islands")
## Increase point size with cex
```

```
plot(pisc$productivity, pisc$biomass, col=pisc$state, pch=2,cex=1.5, xlab="Productivity (Chl-a)",
     ylab="Biomass (grams)", main="Piscivore biomass at Hawaiian islands")
### Change symbol shape with pch
```

What is R using to code the colours? We need a legend.

```
levels(pisc$state) ### R uses the number of factor levels to code colours
?legend ## How do we describe the legend?
legend(0.07, 60000, legend=c('Disturbed', "Remote"), col=c(1,2), pch=1)
```

We can also add a regression line to this relationship.

```
prod_mod<-lm(biomass ~ productivity, pisc)
abline(prod_mod)
```

The plot does suggest that the disturbance state is very important - look at red vs. black points. Let's present the two plots together.

4. Multipanel plots- par()

We use par() to get the graphical parameters - essentially, the formatting of your plot.

```
par(mfrow=c(2,1)) ### Multipanel, 2 rows by 1 column
boxplot(pisc$state, pisc$biomass, col=c(0,2))
## use the same colours as our legend
```

```
plot(pisc$productivity, pisc$biomass, col=pisc$state, pch=1,cex=1.5, xlab="Productivity (Chl-a)",
     ylab="Biomass (grams)", main="Piscivore biomass at Hawaiian islands")
legend(0.07, 60000, legend=c('Disturbed', "Remote"), col=c(1,2), pch=1)
```

Anything else relevant here? Let's look at all of our explanatory variables.

```
par(mfrow=c(2,2))
```

```

plot(pisc$SST, pisc$biomass,col=pisc$state, xlab="SST", ylab="Biomass (grams)")

plot(pisc$richness, pisc$biomass,col=pisc$state,
     xlab="Species richness", ylab="Biomass (grams)")

plot(pisc$log_population, pisc$biomass, col=pisc$state, xlab="Log population density",
     ylab="Biomass (grams)")

plot(pisc$productivity, pisc$biomass, col=pisc$state, pch=1,cex=1,
     xlab="Productivity (Chl-a)", ylab="Biomass (grams)")
legend(0.07, 60000, legend=c('Disturbed', "Remote"),
     col=c(1,2), pch=1, cex=0.5)
## make legend smaller to fit onto multipanel plot (using cex())

```

Play around with the dataset - there are lots of ways to visualise these data. Think of the different trophic groups, species richness, environmental variables, and human population size. Also have a look at the Isle Royale dataset - what can you plot for wolf and moose population sizes?

Useful resources:

- <http://www.statmethods.net/graphs/>
- <http://www.r-bloggers.com/how-to-plot-a-graph-in-r/>

Useful packages:

- "ggplot2" is a popular R library for quick, attractive plot types (<http://docs.ggplot2.org/current/index.html>)
- "lattice" is another multipanel plot library (but the visuals are pretty terrible)