ETC5512: Instructions for Lab 6

Patricia Menéndez

Reading libraries

```r
library(tidyverse)
library(leaflet)
library(ggmap)
library(readr)
library(mapview)
library(viridis)
library(rgdal)
library(lubridate)
```

Lab Objectives

- Explore LTMP data measurements for hard coral cover and algae cover.
- Visualize variability in mean estimates.
- Run a simple regression model.
- Practice your data wrangling skills.

Exercise 2: Create a report and read the data

- Create an Rstudio project.
- Inside your Rstudio project folder create a reproducible Rmd file called *tutorial6.Rmd*. The file should render into an html file.
- Add the *Data* folder place the LTMP data inside that folder.
- Read the LTMP data.

```r
# LTMP data
ltmp <- read_csv(file = "Data/ltmp_hc_sc_a_by_site/ltmp_hc_sc_a_by_site.csv", col_types = cols())
```

- Add an R code chunk to include the relevant R libraries at the top of the *tutorial6.Rmd* file.
- Prepare the data so that you can create the following figures and discuss the differences between those graphs:

Hint:

```r
ltmp %>% dplyr::filter(GROUP_CODE == "Hard Coral",
                       SECTOR == XX
                       ) %>%
                mutate(REEF_NAME = as.factor(REEF_NAME),
                Year = year(XX)) %>%
                group_by(SECTOR, SHELF, Year) %>%
  summarize_at(vars(XX), list(covermean = XX, coversd = XX)) %>%
  mutate(samplesize = n()) %>%
```
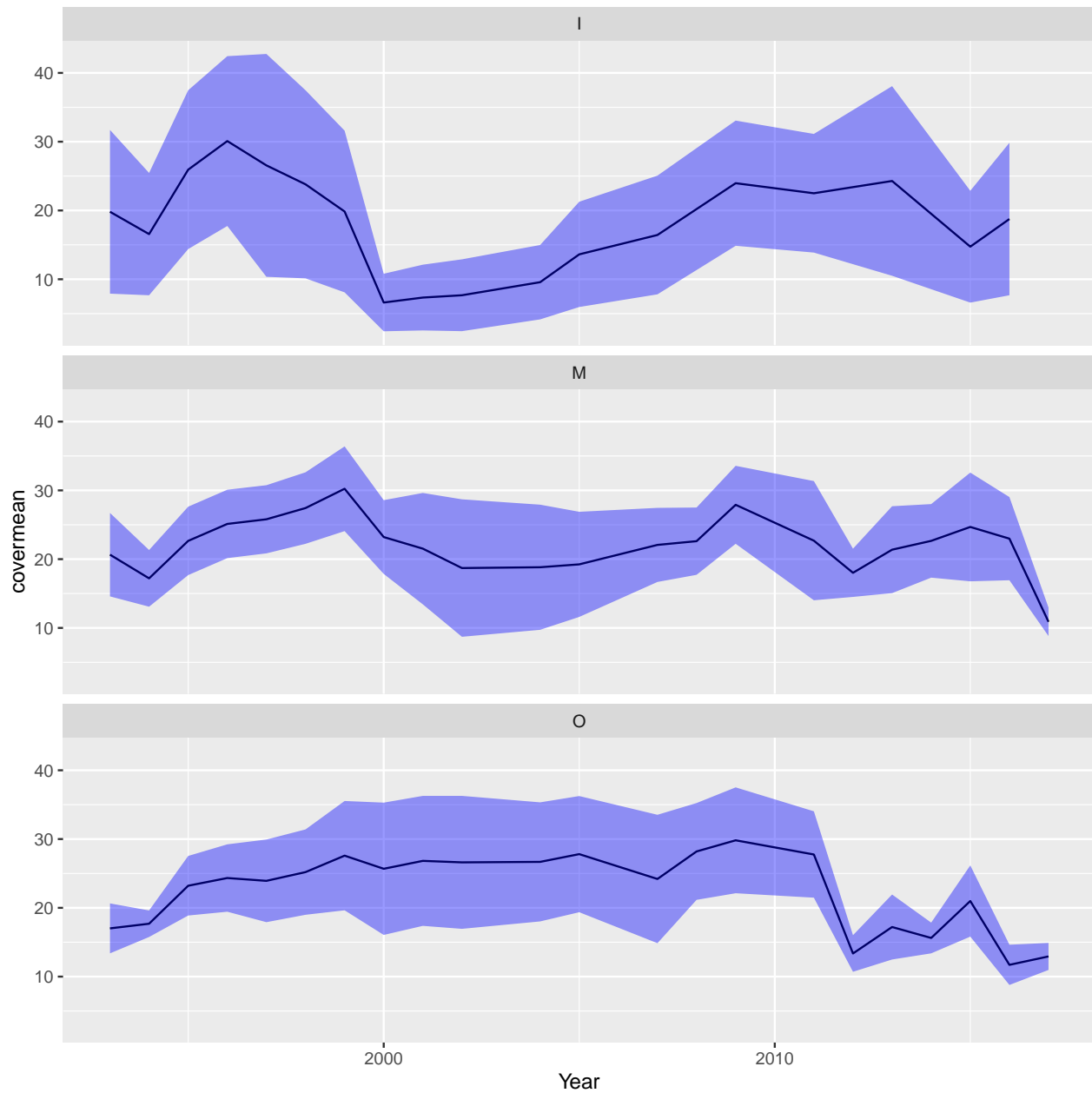
```
mutate(upper = covermean + coversd/sqrt(samplesize),
       lower = covermean - coversd/sqrt(samplesize))
```
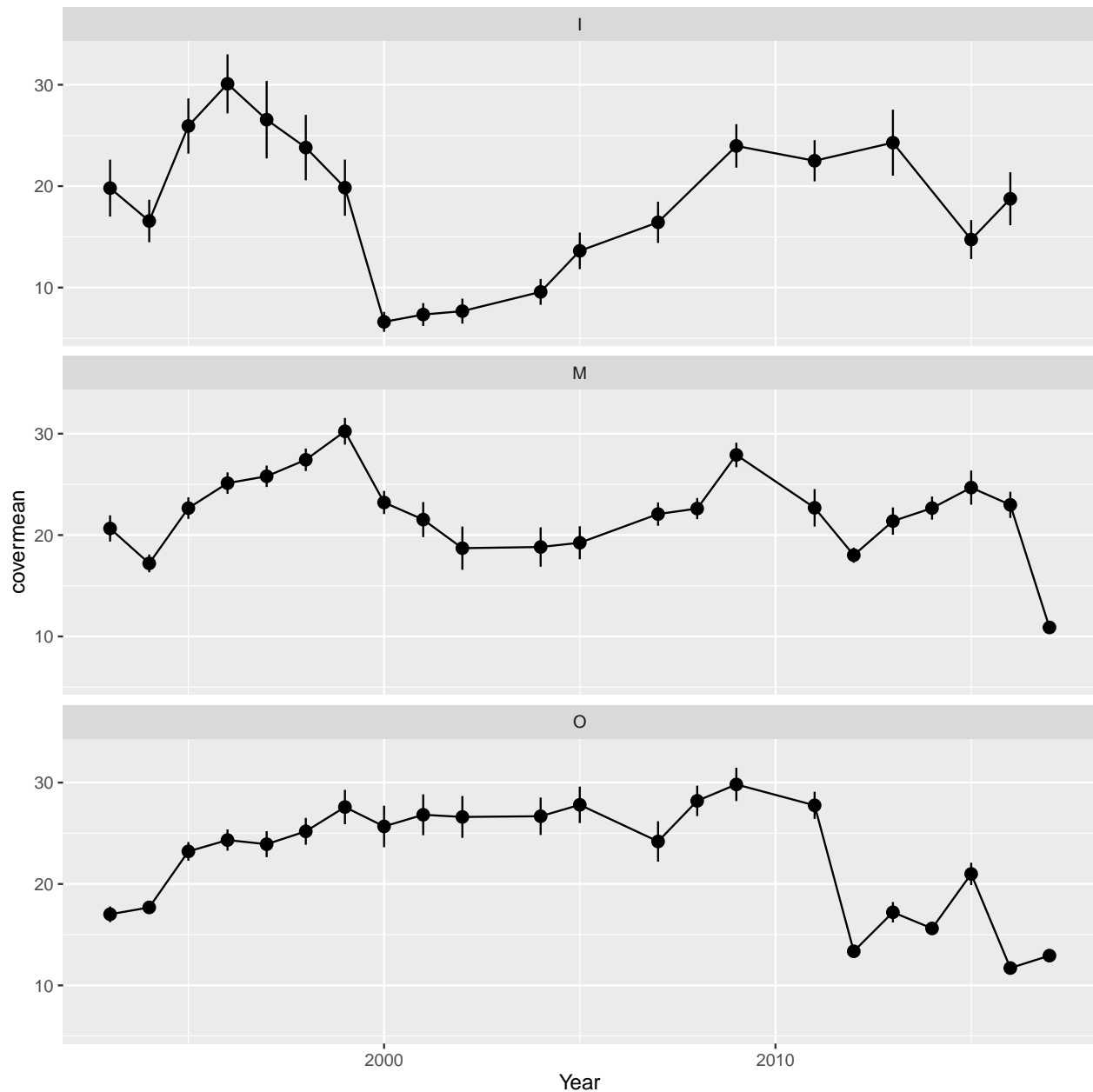

Hint:

```
ltmp %>% dplyr::filter(GROUP_CODE == XX,
                       SECTOR == "CA") %>%
mutate(REEF_NAME = as.factor(REEF_NAME),
       Year = year(SAMPLE_DATE)) %>%
group_by(XX, XX, XX) %>%
summarize_at(vars(COVER), list(covermean = mean, coversd = sd)) %>%
mutate(upper = covermean + XX,
       lower = covermean - XX)
```

Exercise 3: Data wrangling to create summary variables

- Filter the LTMP data set using the variables *GROUP_CODE* and *REEF_NAME* so that you only have values for *Hard Coral* and *Algae* for the *LOW ISLANDS REEF*.
- Create a new variable called *Year* and extract the *year* value from the *SAMPLE_DATE* variable.
- Group the data set by *Year* and *GROUP_CODE* and compute for those groups the mean and stand deviation for the *COVER* variable. Your data should look like this:

```
## # A tibble: 6 x 4
## # Groups:   Year [3]
##    Year GROUP_CODE covermean coversd
##   <dbl> <chr>          <dbl>   <dbl>
## 1  1994 Algae           35.6    3.16
## 2  1994 Hard Coral      24.2    3.85
```

```
## 3  1995 Algae          27.8    5.04
## 4  1995 Hard Coral      30.6    4.06
## 5  1996 Algae          30.2    1.97
## 6  1996 Hard Coral      36.3    4.08
```

- Convert your data into wide format so that you have two new columns where *Hard Coral* and *Algae* cover values are stored. Your data should look like this:

```
## # A tibble: 6 x 3
## # Groups:   Year [6]
##     Year Algae `Hard Coral`
##    <dbl> <dbl>        <dbl>
## 1  1993  34.7         30.2
## 2  1995  32.0         34.7
## 3  1996  35.4         39.2
## 4  1997  37.8         35.4
## 5  1998  43.1         30.3
## 6  1999  49.3         21.3
```

- Plot your newly created data to produce the following figure:
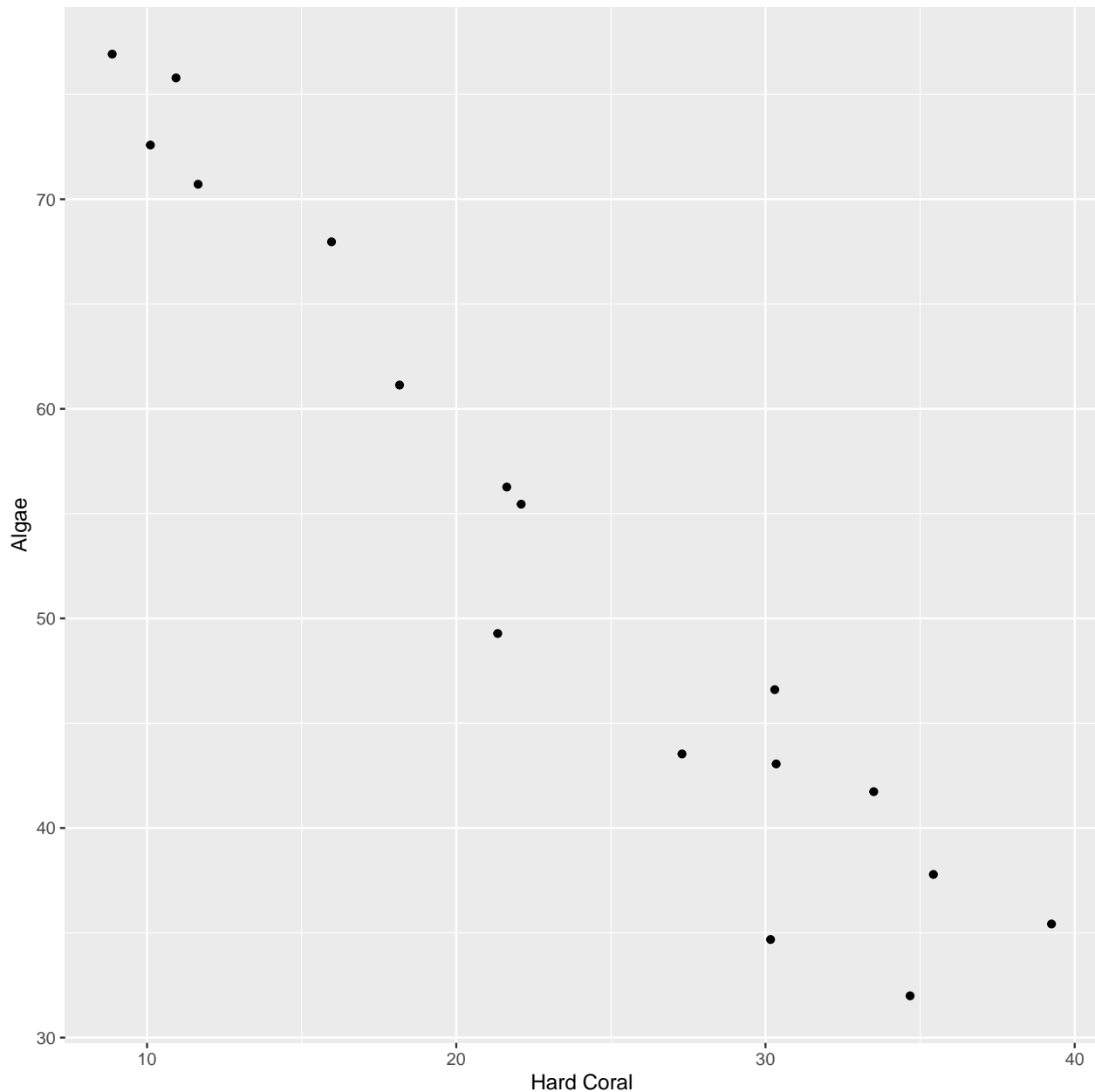
Exercise 4: Practicing liner regression to understand the relationship between hard coral cover and algae abundance.

- Using the data that you created in Exercise 3 (wide format data).

- Run a linear regression model to understand the relationship between hard coral cover and algae.

- Add two new variables into the data set to store your predicted and residuals values from the regression.

- Create the following figure and change the colours:

```
ggplot(regdataWF, aes(x = `Hard Coral`, y = Algae)) +
  geom_smooth(method = "lm", se = FALSE, color = "lightblue") +
  geom_segment(aes(xend = `Hard Coral`, yend = predicted), alpha = .2) +
```

```
geom_point(aes(color = abs(residuals), size = abs(residuals))) +
scale_color_continuous(low = "#D19494", high = "#D92626") +
guides(color = FALSE, size = FALSE) +  # Size legend also removed
geom_point(aes(y = predicted), shape = 1) +
theme_bw() +
ggtitle("Linear regression model")
```


Linear regression model