Coding together week 2 - Data wrangling I

This lesson covers:

An intro to dplyr:

Transforming tables: mutate() adds new variables that are functions of existing variables select() picks variables based on their names. filter() picks cases based on their values. summarise() reduces multiple values down to a single summary. arrange() changes the ordering of the rows.

Vectors, types, indexing and data frames

1. Make a character vector of three names
2. Make a numeric vector of three numbers
3. Make a factor vector of three fruit
4. Combine into a data frame.

Formative exercises

arrange:

Use arrange to find the earliest and latest years in the dataset:

```
surveys %>% arrange(year)

surveys %>% arrange(desc(year))
```

filter:

Use filter to

Filter observations that only occurred on the 9th of March 1986:

```
filter(surveys, month == 3 & day == 9 & year == 1986 & species_id == "NL") %>% View()
```

Use arrange and filter to find the heaviest Kangeroo rat, Krats are DM,DO and DS.

```
surveys %>% filter(species_id == "DM" | species_id == "DO" | species_id == "DS") %>%
arrange(desc(weight))
```

select:

Create an object called surveys small that filters weight less than 5 and selects the columns species_id, sex and weight. Use the pipe.

```
surveys_sml <- surveys %>%
  filter(weight < 5) %>%
  select(species_id, sex, weight)

surveys_sml
```

mutate:

Use mutate to first create a weight_kg variable and then create another variable weight_lb using weight_kg multiplied by 2.2. You don't need to create an object.

```
surveys %>%
  mutate(weight_kg = weight / 1000,
         weight_lb = weight_kg * 2.2)
```

summarise:

Use filter with is.na() to remove the NA values from the weight variable, the use summarise to create mean_weight and min_weight variables, using mean() and min() functions.

```
surveys %>%
  filter(!is.na(weight)) %>%
  summarize(mean_weight = mean(weight),
  min_weight = min(weight))
```

group_by:

Group the surveys data by sex and then use summarise with the n() function to create a count variable, that gives the number of male and female animals.

```
surveys %>%
    group_by(sex) %>%
    summarise(count = n())
```

Use surveys_mutated to group_by rodent_type and then summarrise, we should have 8 species of 2 types.

```
surveys_mutated %>% group_by(rodent_type) %>% summarise()
```

Summative exercise

By semester from 1980 to 2000.

```
surveys %>%
filter(plot_id %in% exp_plots,
year >= 1980 & year <= 2000) %>%
mutate(rodent_type = case_when(
         species_id == "DM" ~ "Kangaroo Rat",
         species_id == "DO" ~ "Kangaroo Rat",
         species_id == "DS" ~ "Kangaroo Rat",
         species_id == "PP" ~ "Granivore",
         species_id == "PF" ~ "Granivore",
         species_id == "PE" ~ "Granivore",
         species_id == "PM" ~ "Granivore",
         species_id == "RM" ~ "Granivore",
         TRUE ~ "Other"),
         date = make_date(day = day, month = month, year = year),
               semester = semester(date, with_year = TRUE)) %>%
               group_by(rodent_type,plot_type,semester) %>%
```

```
      summarise(captures = n()/2) %>%
      filter(rodent_type != "Other") %>%
      ggplot(aes(x=semester,y=captures,colour=rodent_type)) +
geom_line() +
geom_point() +
facet_wrap(~ plot_type) +
      theme(legend.position = "bottom") +
      ggtitle("How does excluding Kangeroo Rats effect Granivore populations?",
              subtitle = "Mean half yearly observations")
```