Exploring Similarities and Differences in Ecosubsections

```r
# Import data
spatial <- read_csv("../data/plot_level/plt_spatial.csv")
response <- read_csv("../data/plot_level/plot_response.csv")

# Join data
## Keep only observations in both `spatial` and `response`
dat <- inner_join(spatial, response,
                  by = c("PLT_CN" = "PLT_CN",
                         "INVYR" = "INVYR"))

# Create columns for province, sections, and subsections
dat <- dat %>%
  mutate(
    subsection = ECOSUBCD.x,
    section = str_remove_all(ECOSUBCD.x, "[:lower:]"),
    province = str_sub(section, end = -2)
  )

# Select small subset of columns to work with for this EDA
dat_small <- dat %>%
  select(PLT_CN, INVYR, PLOT.x, LON_PUBLIC.x, LAT_PUBLIC.x, LON_PUBLIC.y, LAT_PUBLIC.y,
         ELEV_PUBLIC.x, ELEV_PUBLIC.y, forgrp, forprob, nlcd11, demLF, evtLF, forbio,
         BALIVE_TPA, CNTLIVE_TPA, BIOLIVE_TPA, VOLNLIVE_TPA, subsection, section, province)

# Remove redundent columns, rename columns for ease of use
dat_small <- dat_small %>%
  select(-LON_PUBLIC.y, -LAT_PUBLIC.y, -ELEV_PUBLIC.y) %>%
  rename(PLOT = PLOT.x,
         LON_PUBLIC = LON_PUBLIC.x,
         LAT_PUBLIC = LAT_PUBLIC.x,
         ELEV_PUBLIC = ELEV_PUBLIC.x)
```

```r
n_subsections <- dat_small %>%
  group_by(section, subsection) %>%
  summarize(n()) %>%
  group_by(section) %>%
  summarize(number_of_subsections = n())
```

```
## `summarise()` regrouping output by 'section' (override with `.groups` argument)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
head(n_subsections)
```

```
## # A tibble: 6 x 2
##   section number_of_subsections
##   <chr>                   <int>
## 1 313A                       19
## 2 313B                        7
## 3 313C                        4
## 4 313D                        5
## 5 315A                        3
## 6 315B                        4
```

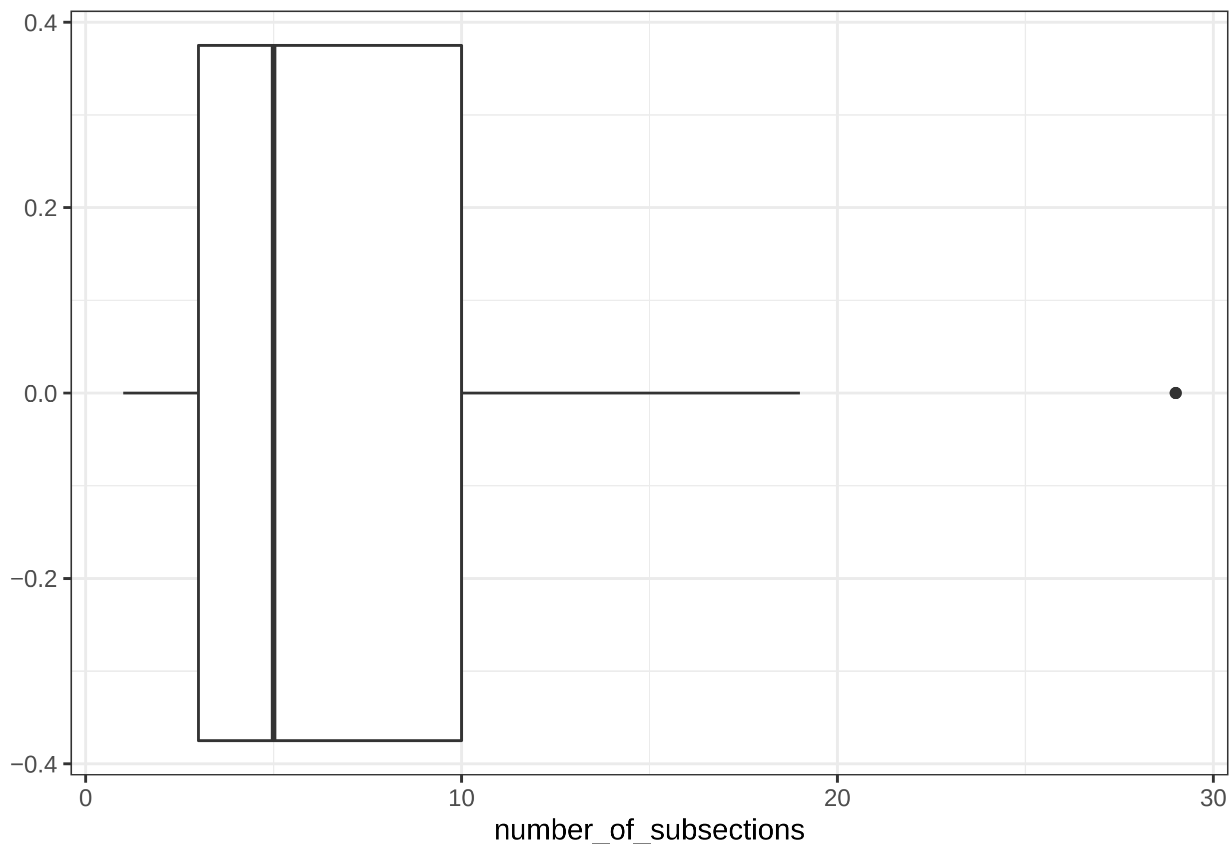
```r
## Does this make sense?
sum(n_subsections$number_of_subsections)
```

```
## [1] 480
```

```r
length(unique(dat_small$subsection))
```

```
## [1] 480
```

```r
## Yes!

## Let's look at this distribution:
ggplot(n_subsections, aes(x = number_of_subsections)) +
  geom_boxplot() +
  theme_bw()
```




```r
mean(n_subsections$number_of_subsections)
```

```
## [1] 7.058824
```

```r
sd(n_subsections$number_of_subsections)
```

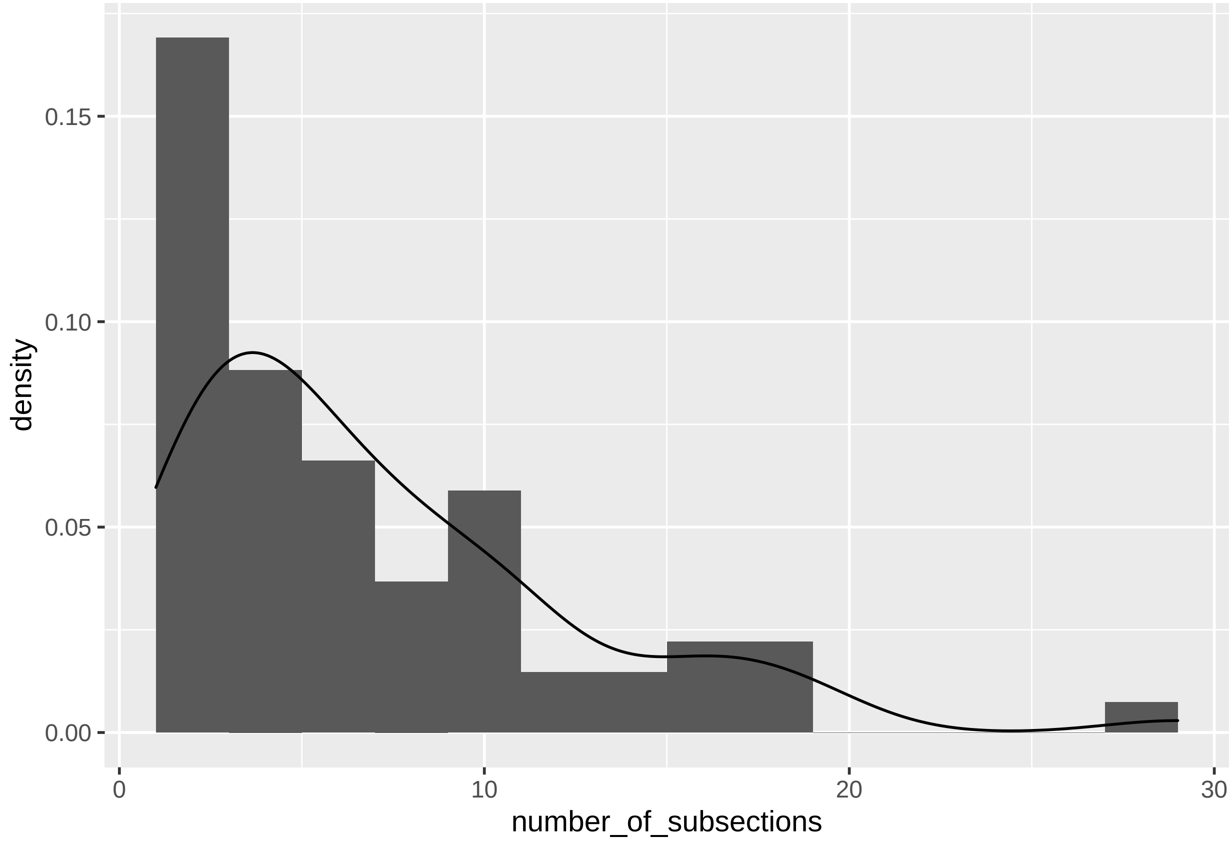
```
## [1] 5.5068
```

```r
median(n_subsections$number_of_subsections)
```

```
## [1] 5
```

```r
ggplot(n_subsections, aes(x = number_of_subsections)) +
  geom_histogram(bins = 15, aes(y = ..density..)) +
```

```r
  geom_density()
```



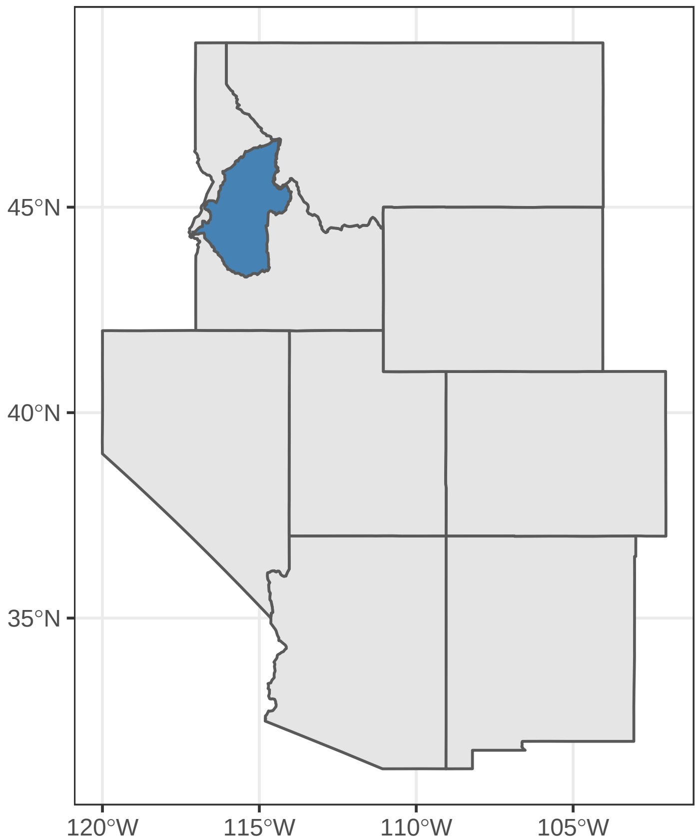
From this, we see that the average number of subsections in a section is about 7, with a median of 5 giving us a right-skewed distribution. There is one outlying section which I will investigate now:

```r
# M332A: Idaho Batholith, "The batholith section is a large, contiguous uplifted area of granitic pluto
library(concaveman)
library(sf)
```

```
## Linking to GEOS 3.7.2, GDAL 2.4.2, PROJ 5.2.0
```

```r
library(USAboundaries)
`%ni%` <- Negate(`%in%`)

interior_west <- c("AZ", "CO", "ID", "MT", "NV", "NM", "UT", "WY")

states <- data.frame(state.abb) %>%
  filter(state.abb %ni% interior_west &
           state.abb %ni% c("AK", "HI")) %>% pull()


m332a_poly <- dat_small %>%
  filter(section == "M332A") %>%
  st_as_sf(coords = c("LON_PUBLIC", "LAT_PUBLIC"), crs = 4326) %>%
  concaveman()

ggplot() +
```

```
  geom_sf(data = us_boundaries(type = "state",
                               states = interior_west)) +
  geom_sf(data = m332a_poly,
          fill = "steelblue") +
  theme_bw()
```




```
# Okay, so it is a big section, but is it way bigger than others?
interior_west_sf <- dat_small %>%
  st_as_sf(coords = c("LON_PUBLIC", "LAT_PUBLIC"),
           crs = 4326) %>%
  concaveman()

total_area <- st_area(interior_west_sf)
m332a_area <- st_area(m332a_poly)

m332a_area / total_area # This is the proportion of total area m332a takes up
```

```
## 0.02157692 [1]
```

```
1 / length(unique(dat_small$section)) # This is the proportion of total area an "average" section would
```

```
## [1] 0.01470588
```

```
# M332A takes up more area than average but not *way* more. This means that it likely has some small su

### this doesn't seem right based on the picture. revisit this with fresh eyes tomorrow.
```

Quantifying Homogeneity in Ecosubsections