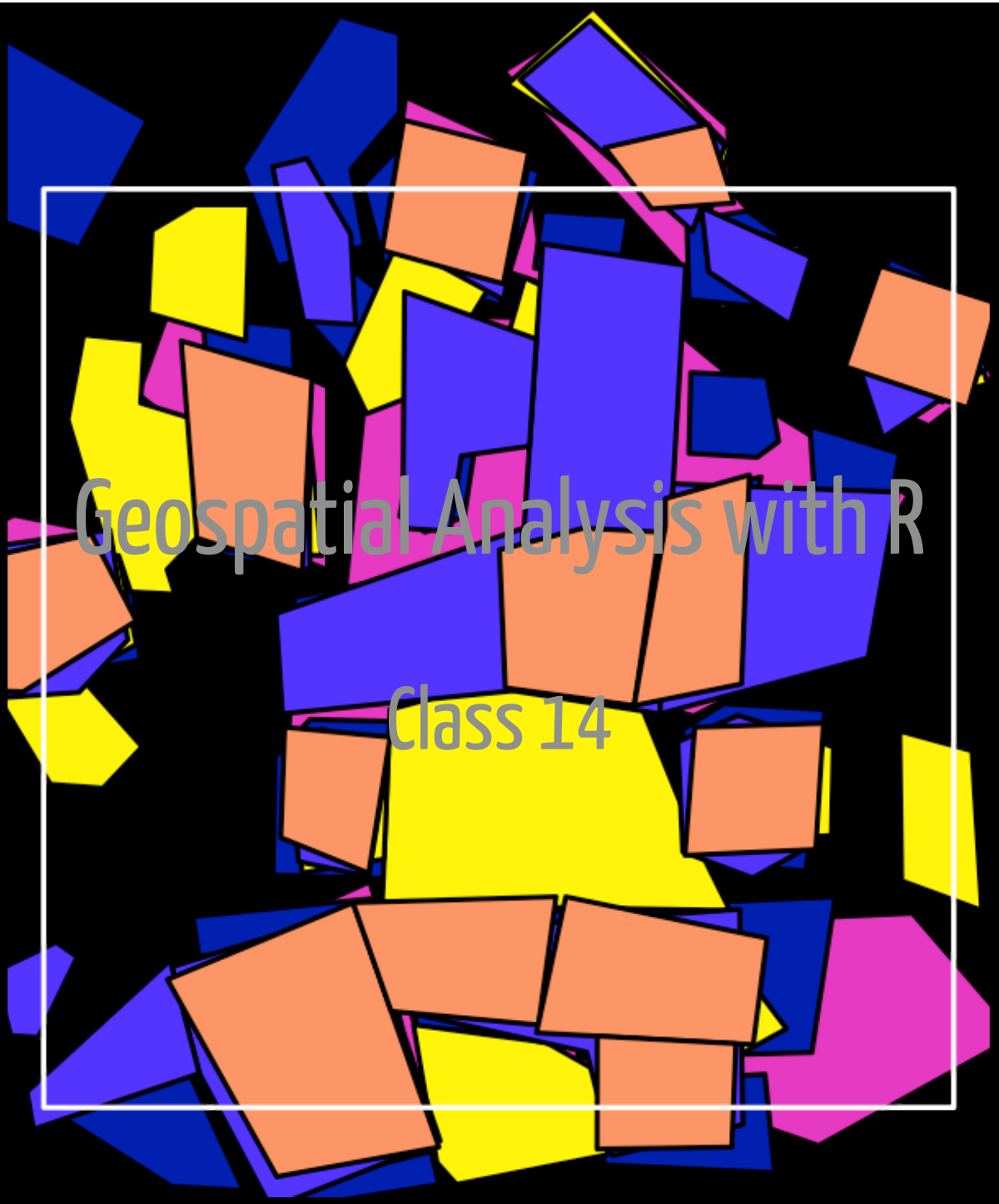
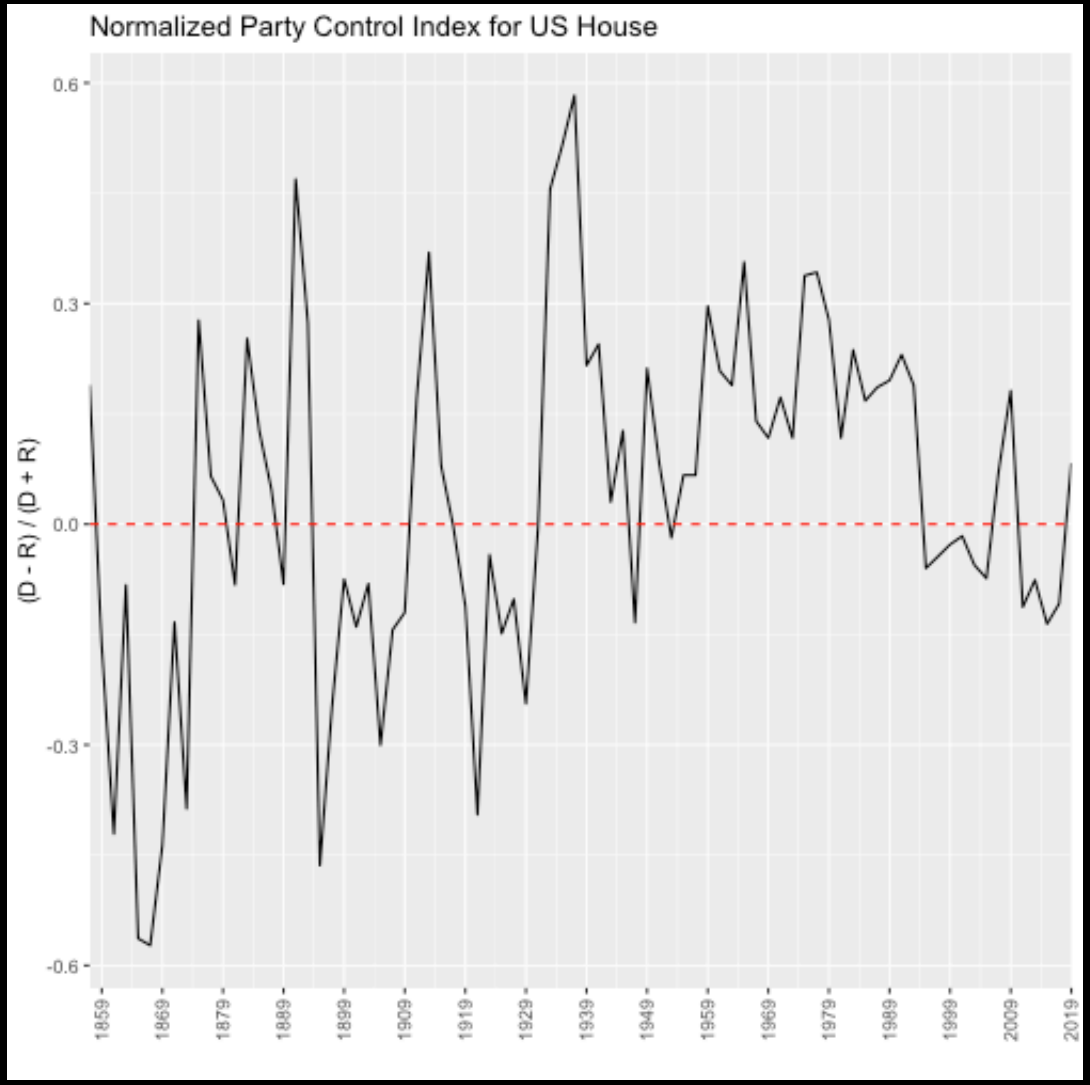

Geospatial Analysis with R
Class 14

# Today

- Visualization basics
- A bit more analytics practice

# Election Results



```r
library(tidyverse)
library(rvest)

URL <- "https://history.house.gov/Institution/Party-Divisions/Party-Divisions/"
webpage <- read_html(URL)
cong <- as_tibble(html_table(webpage)[[1]])
onms <- colnames(cong)
newnms <- c("congress", "seats", "D", "R", "other", "delres")
cong <- cong %>%
  slice((which(grepl("^Republican", `Anti-Administration`))[1] + 1):nrow(.)) %>%
  filter(`Congress (Years)` != colnames(cong)[1]) %>%
  mutate(year = gsub("(*.*-)|(*.*-)|)|)2", "", `Congress (Years)`)) %>%
  mutate(year = as.numeric(year) - 2) %>%
  rename_at(vars(onms), ~newnms) %>% select(-other, -delres) %>%
  mutate(seats = substr(seats, 1, 3)) %>%
  mutate_at(.vars = vars(seats, D, R), as.numeric) %>%
  mutate(swing = (D - R) / (D + R))
cong %>%
  ggplot() + geom_line(aes(year, swing)) +
  scale_x_continuous(breaks = seq(1859, 2019, 10), expand = c(0, 0)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5)) +
  geom_hline(yintercept = 0, lty = 2, col = "red") +
  xlab("") + ylab("(D - R) / (D + R)") +
  ggtitle("Normalized Party Control Index for US House")
```

# Visualization

- `graphics` versus `ggplot`

```
library(tidyverse)
fs <- dir(system.file("extdata/", package = "geospaar"), pattern = "FAOSTAT",
          full.names = TRUE)
crops <- lapply(fs, readr::read_csv)
crops_df <- do.call(rbind, lapply(crops, function(x) {
  x %>% dplyr::select(Item, Area, Element, Year, Value) %>%
    spread(key = Element, value = Value) %>%
    rename(crop = Item, country = Area, year = Year,
           harv_area = `Area harvested`, prod = Production) %>%
    mutate(harv_km2 = harv_area / 100)
}))
crop_ylds <- crops_df %>% mutate(yields = prod / harv_area) %>%
  mutate(country = ifelse(country == "South Africa", "ZAF", country)) %>%
  mutate(country = ifelse(country == "Zambia", "ZMB", country)) %>%
  mutate(harv_km2 = harv_area / 100)
```

# `graphics`

- Need this because we'll use a lot with `sf` and `raster`
- Key functions:
  - `plot`, `barplot`, `hist`, `boxplot`, `par`
- Other helpers:
  - `axis`, `text`, `mtext`, `legend`

## `graphics` examples

- Scatter
- Histogram
- multi-panel

## Scatter

```
# single line
crop_ylds %>% filter(country == "ZAF" & crop == "Maize") %>%
  plot(yields ~ year, data = ., type = "l")

# multiple lines

# change axes

# multiple panels
```

## Histogram

```
# basic
crop_ylds %>% filter(crop == "Maize") %>%
  pull(yields) %>% hist(., breaks = seq(0, 7, 0.5), main = "Maize yields")

# side-by-side
par(mfrow = c(1, 2))
crop_ylds %>% filter(crop == "Maize" & country == "ZAF") %>%
  pull(yields) %>%
  hist(., breaks = seq(0, 7, 0.5), main = "Maize yields", xlim = c(0, 10))
crop_ylds %>% filter(crop == "Maize" & country == "ZMB") %>%
  pull(yields) %>%
  hist(., breaks = seq(0, 7, 0.5), main = "Maize yields", xlim = c(0, 10))
```

# ggplot examples

**scatters**

```
# basic points
crop_ylds %>% filter(crop == "Maize") %>%
  ggplot() + geom_point(aes(x = year, y = yields))

# multiple lines
crop_ylds %>% filter(crop == "Maize") %>%
  ggplot() + geom_line(aes(x = year, y = yields, color = country)) +
  scale_color_manual(values = c("green", "blue"))

# change axes/theme
crop_ylds %>% filter(crop == "Maize") %>%
  ggplot() + geom_line(aes(x = year, y = yields, color = country)) +
  scale_color_manual(values = c("green", "blue")) +
  scale_x_continuous(breaks = seq(1960, 2020, 5), expand = c(0, 0)) +
  theme(axis.text.x = element_text(angle = 90),
        panel.background = element_blank(),
        axis.line = e
```

## multi-panel

```
# from a single variable, facet_grid, facet_wrap
crop_ylds %>%
  ggplot() + geom_line(aes(x = year, y = yields, color = country)) +
  scale_color_manual(values = c("green", "blue")) +
  facet_grid(cols = vars(crop))

# multiple variables
p1 <- crop_ylds %>% filter(crop == "Maize") %>%
  ggplot() + geom_line(aes(x = year, y = yields, color = country)) +
  scale_color_manual(values = c("green", "blue"))
p2 <- crop_ylds %>% filter(crop == "Maize") %>%
  ggplot() + geom_line(aes(x = year, y = harv_area, color = country)) +
  scale_color_manual(values = c("green", "blue"))

gridExtra::grid.arrange(p1, p2, ncol = 2)
g1 <- cowplot::plot_grid(p1 + theme(legend.position = "none"),
                         p2 + theme(legend.position = "none"))
cowplot::plot_grid(g1, cowplot::get_legend(p1), rel_widths = c(2, 0.2))
```

# Practical - data analysis and plotting

- Calculate the mean and standard deviations of crop yield by crop and country
- Use ggplot to plot the histograms of maize yields by country
- Use ggplot to plot the histograms of yields by country and crop (hint: you need to use `facet_grid` on the *crop* variable)
- Fit a regression model to maize yield, where year is the dependent variable. Use base R's `lm` function
- Plot the linear regression fit between Zambian and South Africa wheat yields, using `ggplot` and `geom_smooth`, with `method = "lm"`

# Practical answers

Buried in the Rmarkdown