ESM 206 Lab 1 Summary

Bren School (UCSB)

Allison Horst

Lab 1 Packages Required

tidyverse

Lab 1 Functions

Here, 'df' is the name of the data frame you're working with.

Functions for initial data exploration

- names(df): variable (column) names
- dim(df): data frame dimensions
- class(df): variable classes
- head(df): first 6 lines of data frame
- tail(df): last 6 lines of data frame
- View(df): view data frame in new tab
- summary(df): value and class summary for each column

Functions for basic data wrangling

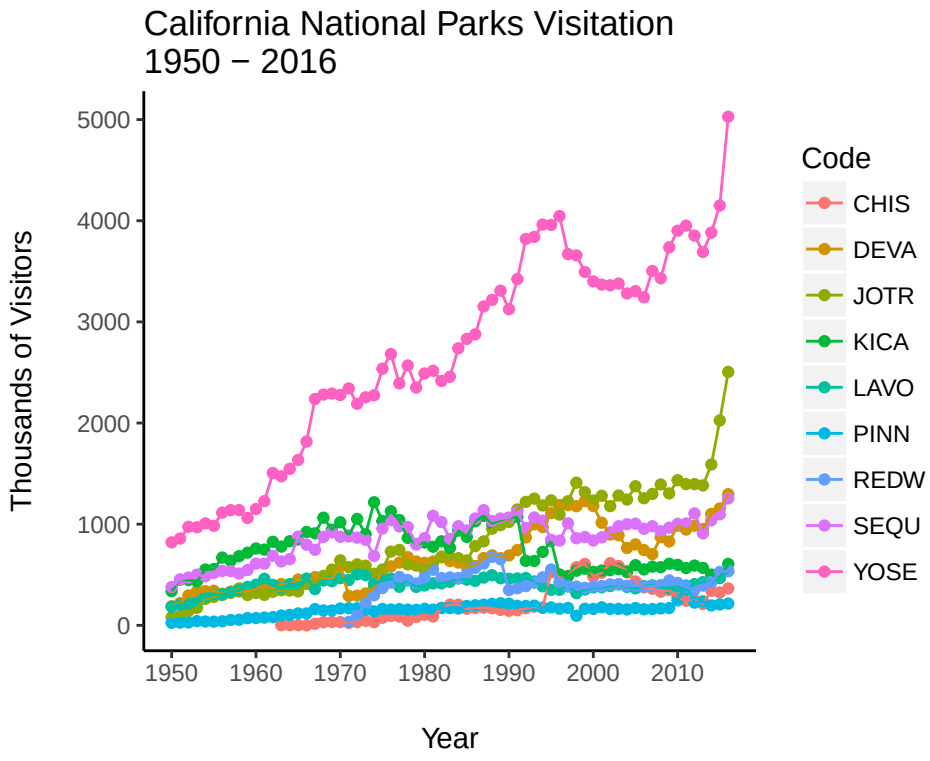
- select(): select only specified columns
- filter(): filter df to retain rows satisfying specified value/character conditions
- arrange(): sort data numerically or alphabetically
- mutate(): add new column(s) based on those existing in data frame

Other useful functions/operators

- df$A: use the dollar sign to refer to a single column
- ?: use a single question mark followed by the function name for R documentation
- ??: use double question mark followed by function name to list all possible R functions
- as.numeric(): convert a variable/vector to 'numeric' class
- as.factor(): convert a variable/vector to 'factor' class
- quartz(): open a new graphics window on a mac
- windows(): open a new graphics window on a PC

Example Graphs

```
ggplot(df5, aes(x = YearRaw, y = kVis)) +
  geom_point(aes(colour = Code)) +
  geom_line(aes(colour = Code)) +
  ggtitle("California National Parks Visitation\n1950 - 2016") +
  xlab("\nYear") +
  ylab("Thousands of Visitors\n") +
  theme(panel.grid.major = element_blank(), panel.grid.minor =
     element_blank(),panel.background = element_blank(), axis.line =
     element_line(colour = "black")) +
  scale_x_continuous(breaks = c(1950,1960,1970,1980,1990,2000,2010,2020))
```



```
ggplot(df5, aes(x = Code, y = kVis)) +
  geom_boxplot(aes(fill = Code)) +
  theme_bw() +
  ggtitle("CA National Park Visitation (1950 - 2016)") +
  xlab("National Park") +
  ylab("Thousands of Visitors") +
  scale_x_discrete(breaks = c("CHIS", "DEVA", "JOTR", "KICA", "LAVO",
                              "PINN", "REDW", "SEQU", "YOSE"),
    labels = c("Channel Islands", "Death Valley", "Joshua Tree", "Kings Canyon",
               "Lassen Volcanic",  "Pinnacles", "Redwoods", "Sequoia", "Yosemite")) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45))
```