graphics with ggplot2

Christin Khan

October 22, 2018

# Learning ggplot2

We’re learning ggplot2 It’s going to be amazing!

## Load packages

# install and load tidyverse which is a package of packages  
#install.packages("tidyverse")  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.4.4

## -- Attaching packages --------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.1 v dplyr 0.7.4  
## v tidyr 0.8.1 v stringr 1.2.0  
## v readr 1.1.1 v forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.4.4

## Warning: package 'tibble' was built under R version 3.4.3

## Warning: package 'tidyr' was built under R version 3.4.4

## Warning: package 'readr' was built under R version 3.4.4

## Warning: package 'purrr' was built under R version 3.4.4

## Warning: package 'dplyr' was built under R version 3.4.3

## Warning: package 'forcats' was built under R version 3.4.4

## -- Conflicts ------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## Load National Parks data

#National Parks in California  
ca <- read.csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/ca.csv")   
  
#Acadia National Park  
acadia <- read.csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/acadia.csv")  
  
#Southeast US National Parks  
se <- read.csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/se.csv")  
  
#2016 Visitation for all Pacific West National Parks  
visit\_16 <- read.csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/visit\_16.csv")  
  
#All Nationally designated sites in Massachusetts  
mass <- read.csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/mass.csv")

# explore data

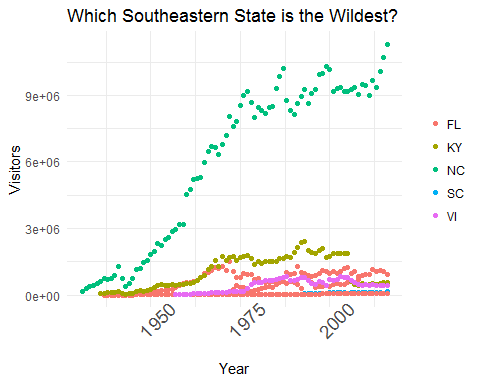
head(se)

## region state code park\_name type visitors year  
## 1 SE FL BISC Biscayne National Park National Park 78147 1972  
## 2 SE FL BISC Biscayne National Park National Park 243100 1973  
## 3 SE FL BISC Biscayne National Park National Park 409700 1974  
## 4 SE FL BISC Biscayne National Park National Park 351500 1975  
## 5 SE FL BISC Biscayne National Park National Park 339100 1976  
## 6 SE FL BISC Biscayne National Park National Park 205900 1977

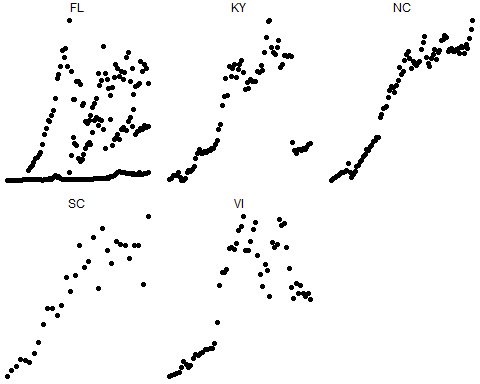
View(se)

# play with data where geom\_point is a scatter plot

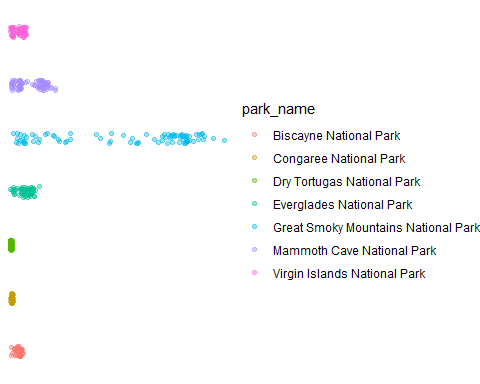
ggplot(data = se) +  
 geom\_point(aes (x = year, y =visitors, color=state)) +  
 theme\_minimal()+  
 xlab("Year") +  
 ylab("Visitors") +  
 ggtitle("Which Southeastern State is the Wildest?") +  
 theme(legend.title = element\_blank(),axis.text.x = element\_text(angle = 45, size =14))



ggplot(data = se) +  
 geom\_point(aes(x = year, y = visitors)) +  
 facet\_wrap(~ state, scales = "free") +  
 theme\_void()



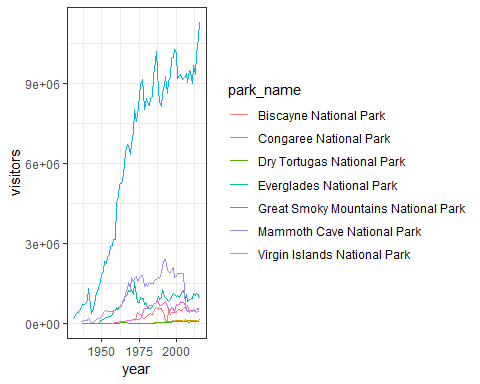
ggplot(data = se) +  
 geom\_jitter(aes(x = park\_name, y = visitors, color = park\_name), alpha = 0.4, width = 0.1) +  
 theme\_void() + coord\_flip()



ggplot(data = se) +  
 geom\_violin(aes(x = park\_name, y = visitors)) +  
 theme\_void()

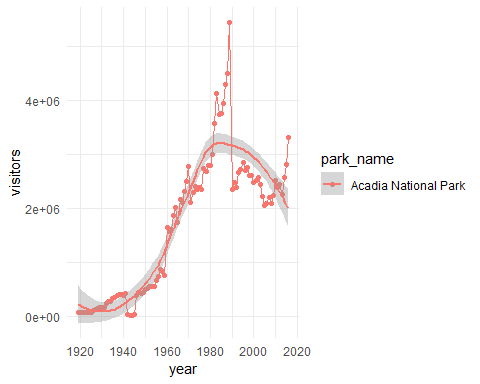


ggplot(se) +  
 geom\_line(aes(x = year, y = visitors, color=park\_name)) +  
 theme\_bw()

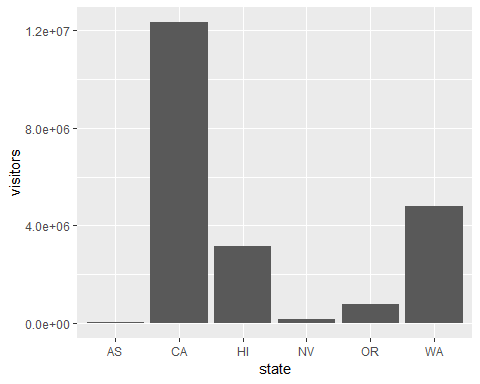
 # multiple geoms on one plot

ggplot(data = acadia, aes(x = year, y = visitors, color=park\_name)) +  
 geom\_point() +  
 geom\_line() +  
 geom\_smooth() +  
 theme\_minimal()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

 # bar charts + how to name and save your ggplot

myplot <- ggplot(data = visit\_16, aes(x = state, y = visitors)) +  
geom\_bar(stat = "identity")  
  
myplot



ggsave("myplot.png", myplot)

## Saving 5 x 4 in image