

# DataViz1-ggplotBasics

Ellie Kuskie

## Setup

```
library(ggplot2)
library(here)
```

here() starts at C:/Users/conno/OneDrive/Documents/UCSC Undergrad/DataScience4EEB/DataViz1-g

```
# Read in the data and store as "ci_np" data object
ci_np <- read.csv("ci_np.csv")

# display the first 6 rows
head(ci_np)
```

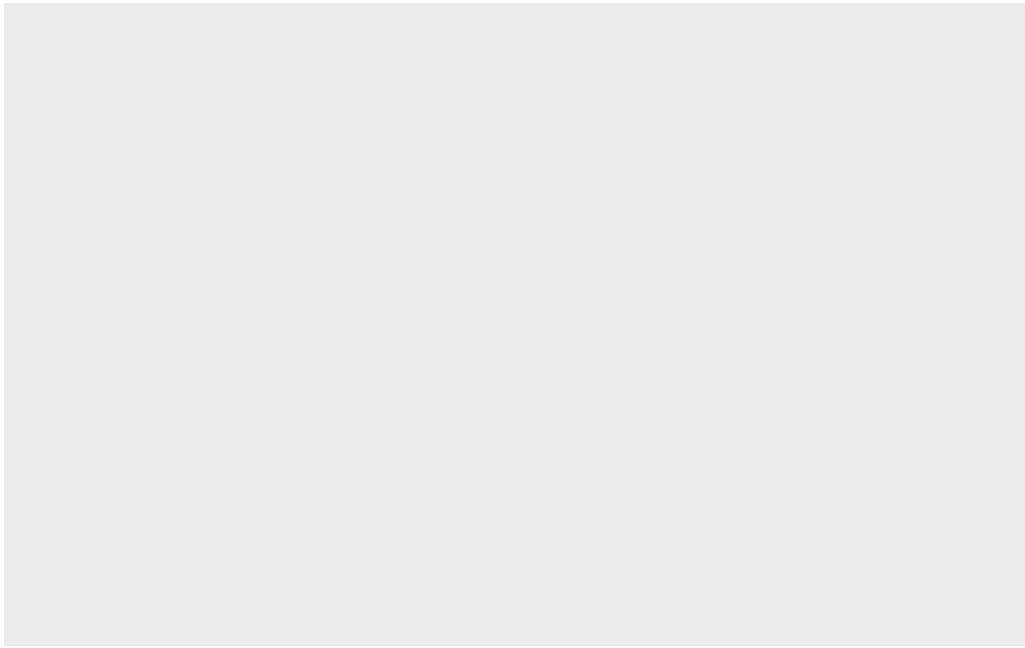
	region	state	code	park_name	type	visitors	year
1	PW	CA	CHIS	Channel Islands National Park	National Park	1200	1963
2	PW	CA	CHIS	Channel Islands National Park	National Park	1500	1964
3	PW	CA	CHIS	Channel Islands National Park	National Park	1600	1965
4	PW	CA	CHIS	Channel Islands National Park	National Park	300	1966
5	PW	CA	CHIS	Channel Islands National Park	National Park	15700	1967
6	PW	CA	CHIS	Channel Islands National Park	National Park	31000	1968

## Q1)

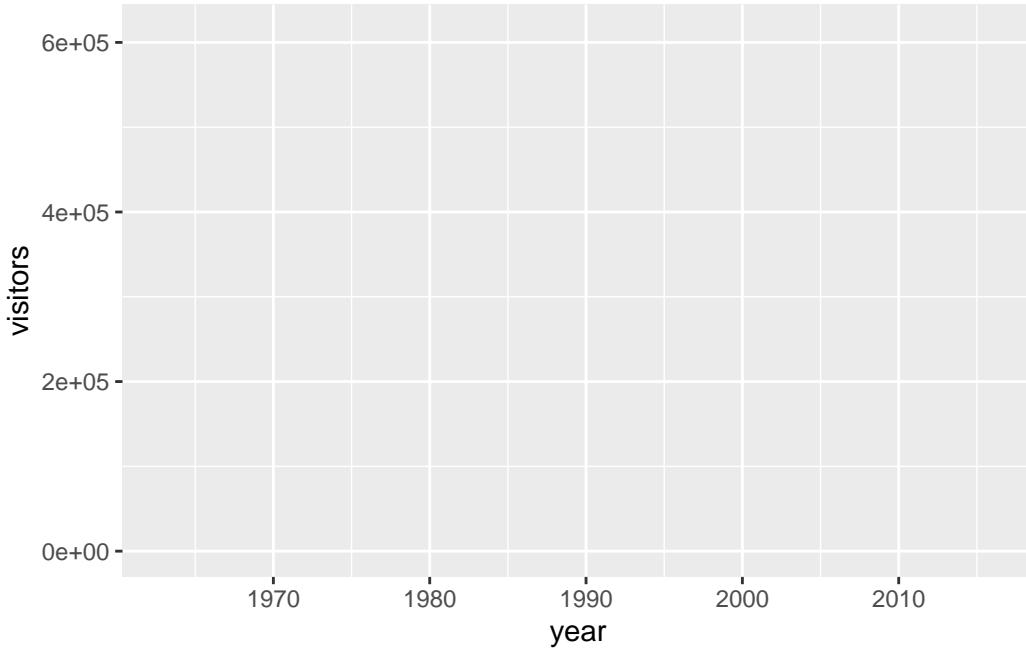
```
# got column names from ci_np dataset
colnames(ci_np)
```

```
[1] "region"      "state"       "code"        "park_name"    "type"        "visitors"
[7] "year"
```

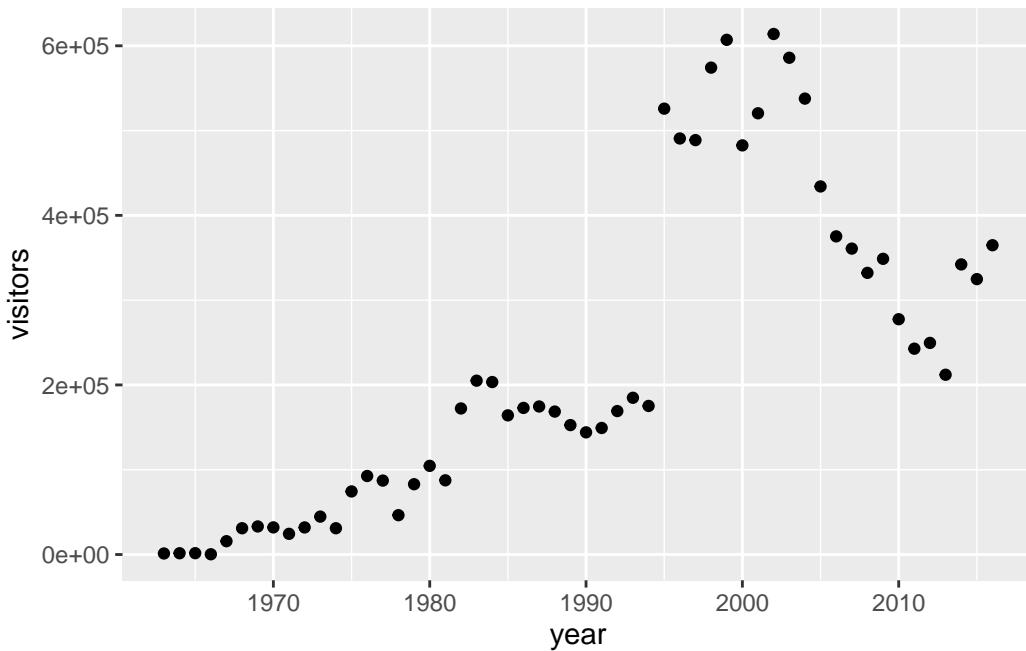
```
# created blank plot (no axes or points)
ggplot(data = ci_np)
```



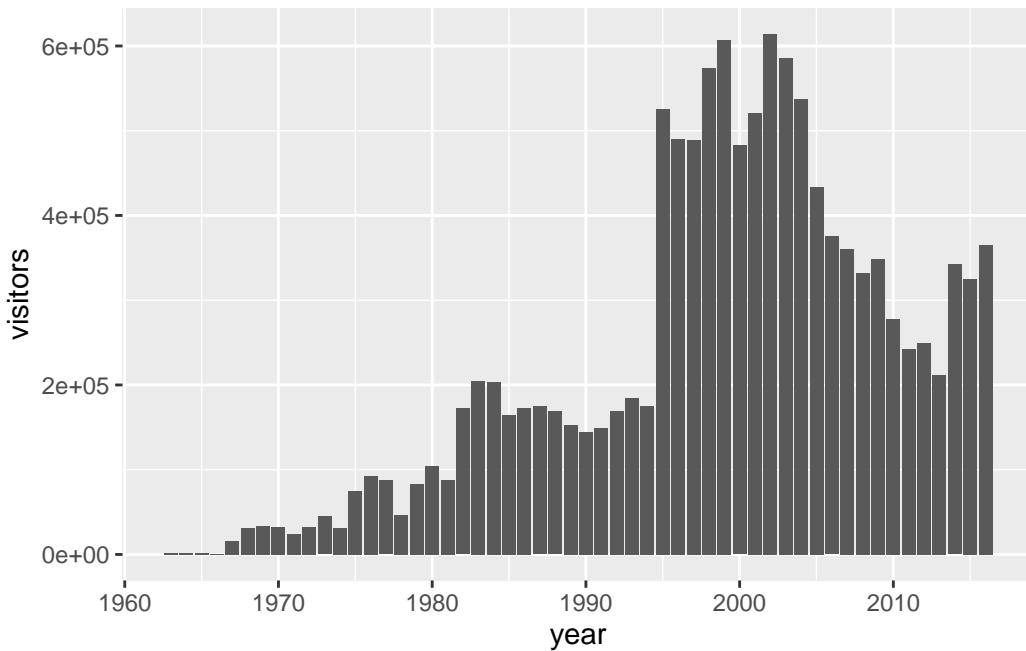
```
# made plot with axes but still no points
ggplot(data = ci_np, mapping = aes(x = year, y = visitors))
```



```
# added points to graph
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point()
```

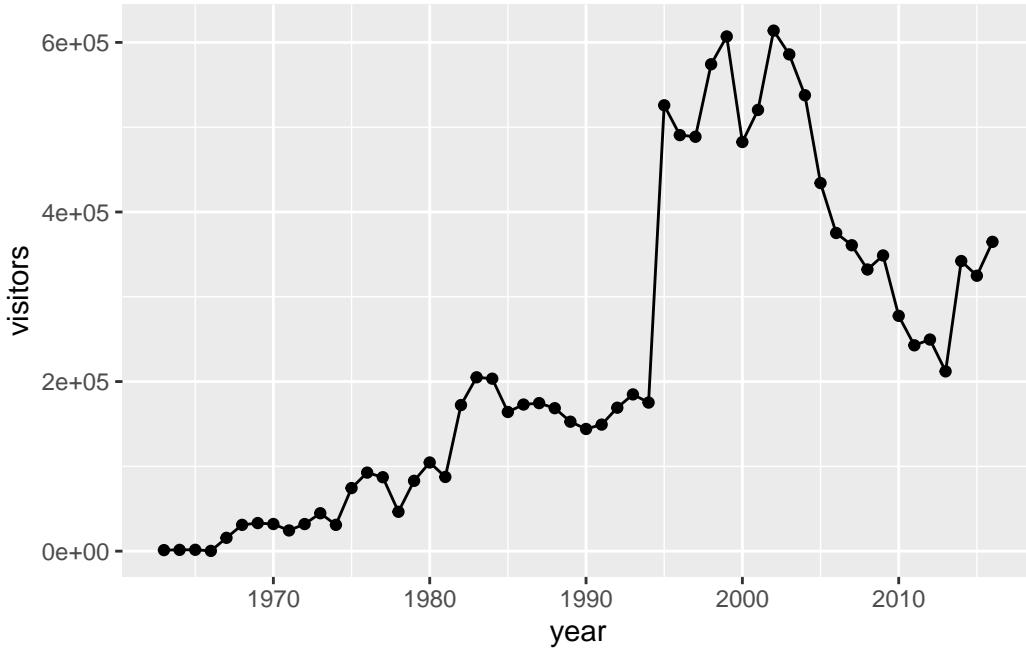


```
# changed graph to column chart
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_col()
```



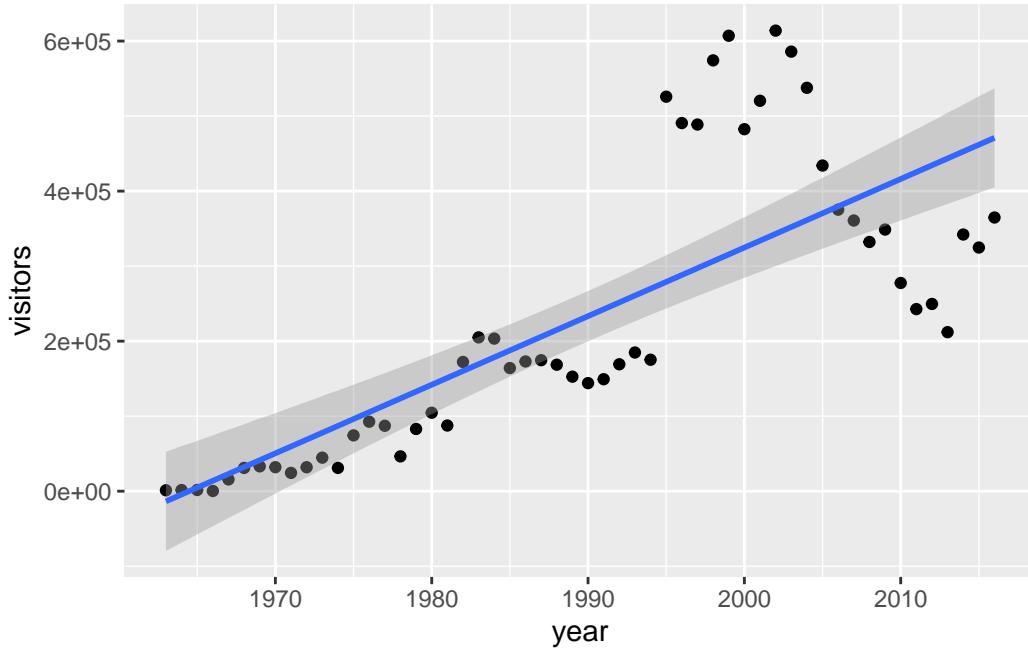
## Q2)

```
# made into line graph
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point() +
  geom_line()
```



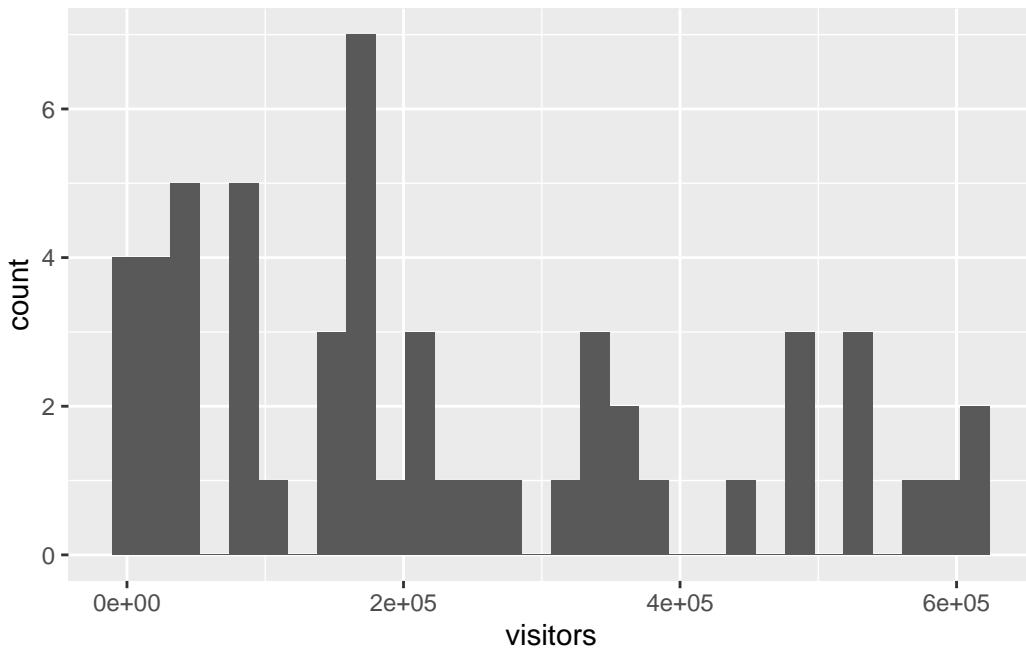
```
# made smooth trendline instead of just connecting dots
# then made it linear using method = lm
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point() +
  geom_smooth(method = lm)

`geom_smooth()` using formula = 'y ~ x'
```

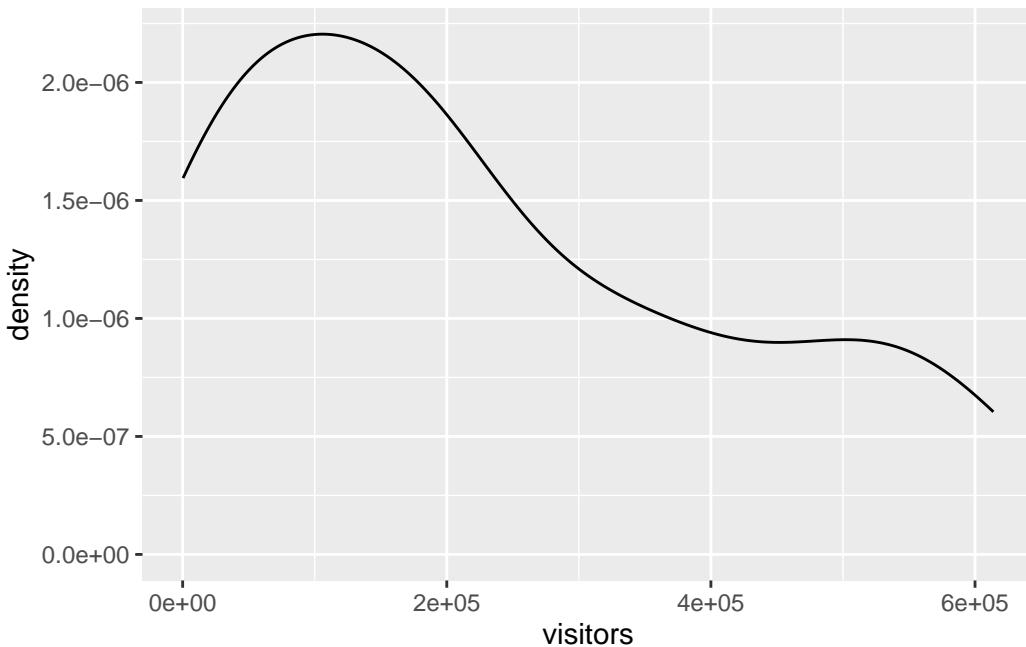


```
# created histogram
ggplot(data = ci_np, mapping = aes(x = visitors)) +
  geom_histogram()
```

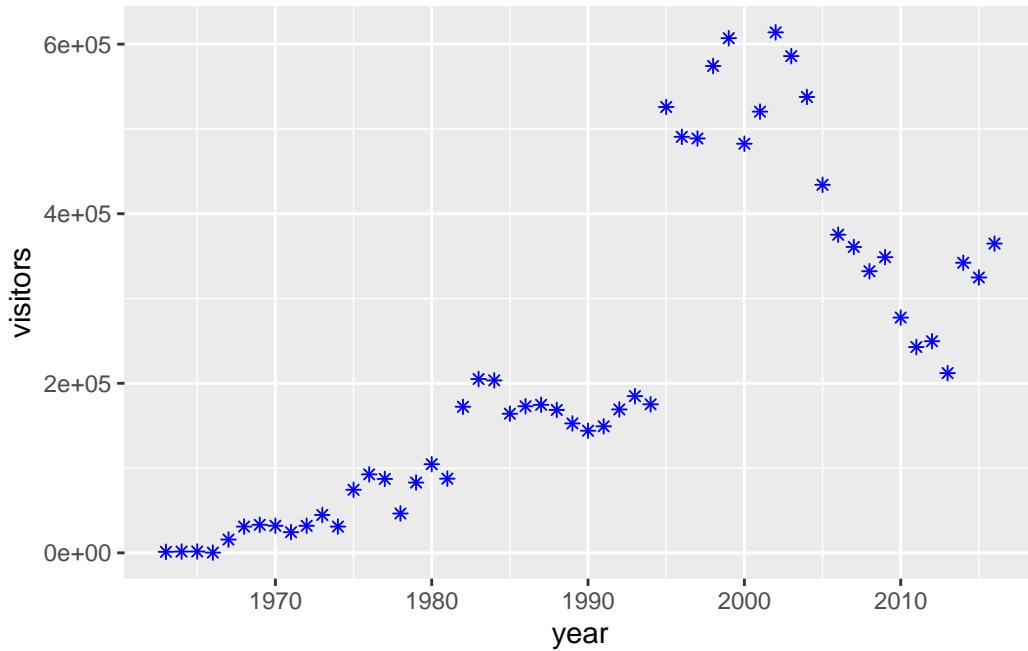
```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# created density line instead of histogram
ggplot(data = ci_np, mapping = aes(x = visitors)) +
  geom_density()
```

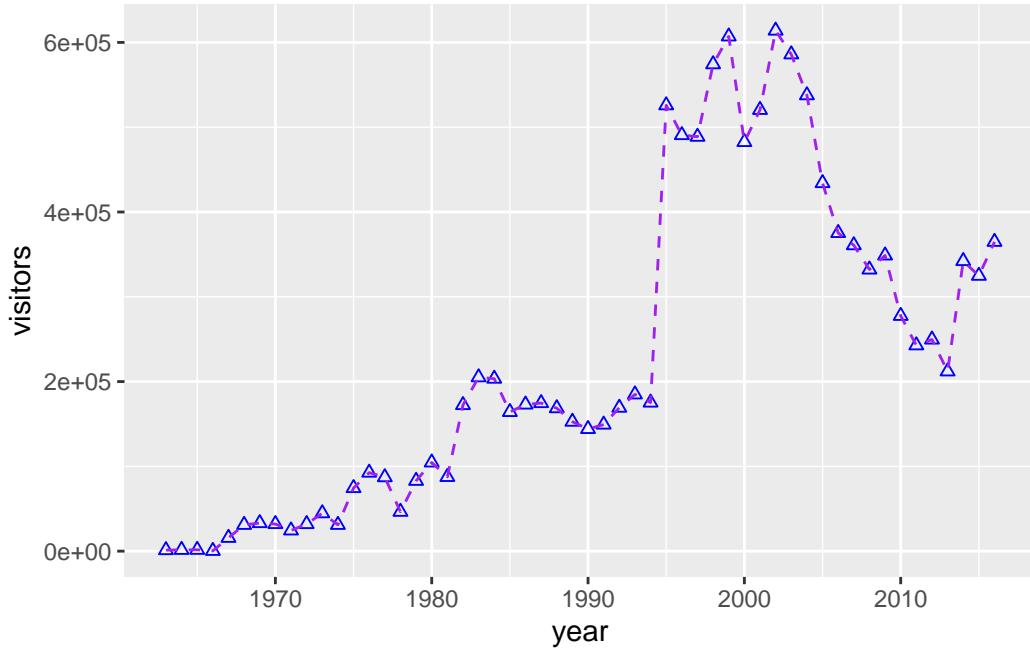


```
# made points on graph blue and differently shaped
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point(shape = 8, color = "blue")
```

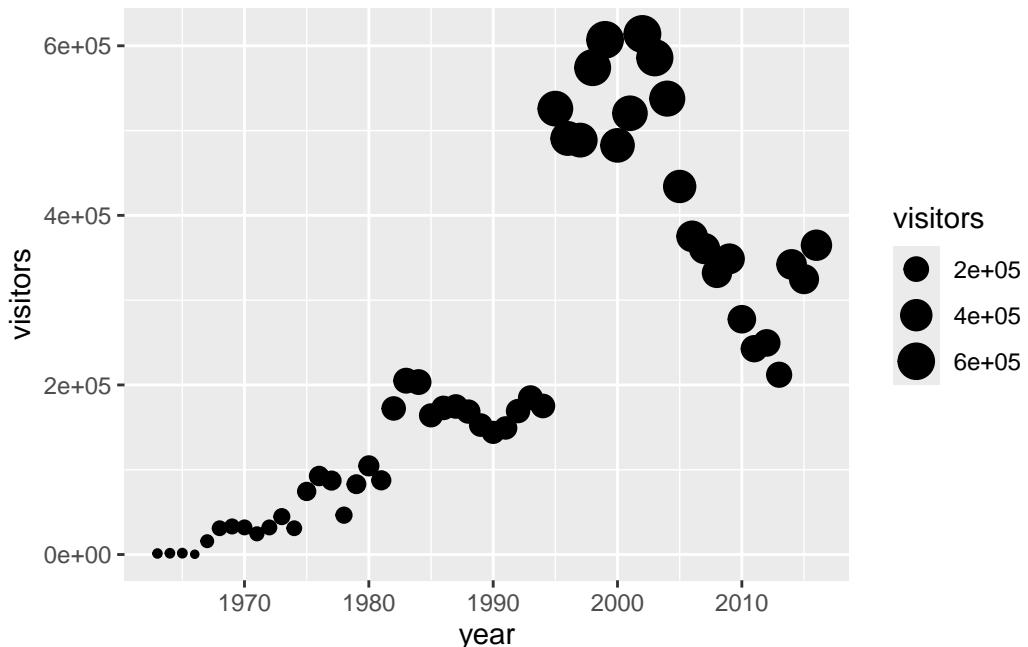


### Q3)

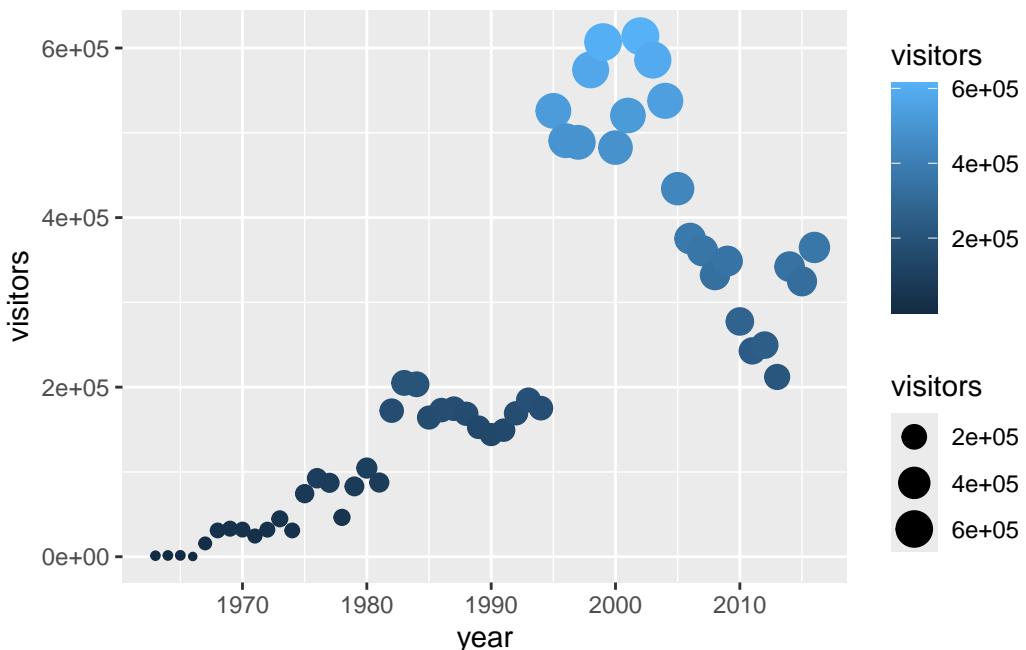
```
# made line plot but changed line color and type
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point(color = "blue",
             shape = 2) +
  geom_line(color = "purple",
            linetype = "dashed")
```



```
# changed point size based on visitor number
ggplot(ci_np, aes(x = year, y = visitors)) +
  geom_point(aes(size = visitors))
```



```
# also color coded the number of visitors
ggplot(ci_np, aes(x = year, y = visitors)) +
  geom_point(aes(size = visitors,
                 color = visitors))
```



#### Q4)

```
# read new csv file
ca_np <- read.csv("ca_np.csv")

# Fetch the column only - not informative! Too many values!
ca_np$park_name
```

```
[1] "Channel Islands National Park" "Channel Islands National Park"
[3] "Channel Islands National Park" "Channel Islands National Park"
[5] "Channel Islands National Park" "Channel Islands National Park"
[7] "Channel Islands National Park" "Channel Islands National Park"
[9] "Channel Islands National Park" "Channel Islands National Park"
[11] "Channel Islands National Park" "Channel Islands National Park"
[13] "Channel Islands National Park" "Channel Islands National Park"
```





[187] "Joshua Tree National Park"  
[189] "Joshua Tree National Park"  
[191] "Joshua Tree National Park"  
[193] "Joshua Tree National Park"  
[195] "Joshua Tree National Park"  
[197] "Joshua Tree National Park"  
[199] "Joshua Tree National Park"  
[201] "Joshua Tree National Park"  
[203] "Joshua Tree National Park"  
[205] "Joshua Tree National Park"  
[207] "Joshua Tree National Park"  
[209] "Joshua Tree National Park"  
[211] "Joshua Tree National Park"  
[213] "Joshua Tree National Park"  
[215] "Kings Canyon National Park"  
[217] "Kings Canyon National Park"  
[219] "Kings Canyon National Park"  
[221] "Kings Canyon National Park"  
[223] "Kings Canyon National Park"  
[225] "Kings Canyon National Park"  
[227] "Kings Canyon National Park"  
[229] "Kings Canyon National Park"  
[231] "Kings Canyon National Park"  
[233] "Kings Canyon National Park"  
[235] "Kings Canyon National Park"  
[237] "Kings Canyon National Park"  
[239] "Kings Canyon National Park"  
[241] "Kings Canyon National Park"  
[243] "Kings Canyon National Park"  
[245] "Kings Canyon National Park"  
[247] "Kings Canyon National Park"  
[249] "Kings Canyon National Park"  
[251] "Kings Canyon National Park"  
[253] "Kings Canyon National Park"  
[255] "Kings Canyon National Park"  
[257] "Kings Canyon National Park"  
[259] "Kings Canyon National Park"  
[261] "Kings Canyon National Park"  
[263] "Kings Canyon National Park"  
[265] "Kings Canyon National Park"  
[267] "Kings Canyon National Park"  
[269] "Kings Canyon National Park"  
[271] "Kings Canyon National Park"







[531] "Redwood National Park"  
[533] "Redwood National Park"  
[535] "Redwood National Park"  
[537] "Redwood National Park"  
[539] "Redwood National Park"  
[541] "Redwood National Park"  
[543] "Redwood National Park"  
[545] "Redwood National Park"  
[547] "Redwood National Park"  
[549] "Redwood National Park"  
[551] "Redwood National Park"  
[553] "Redwood National Park"  
[555] "Redwood National Park"  
[557] "Redwood National Park"  
[559] "Redwood National Park"  
[561] "Redwood National Park"  
[563] "Redwood National Park"  
[565] "Redwood National Park"  
[567] "Redwood National Park"  
[569] "Sequoia National Park"  
[571] "Sequoia National Park"  
[573] "Sequoia National Park"  
[575] "Sequoia National Park"  
[577] "Sequoia National Park"  
[579] "Sequoia National Park"  
[581] "Sequoia National Park"  
[583] "Sequoia National Park"  
[585] "Sequoia National Park"  
[587] "Sequoia National Park"  
[589] "Sequoia National Park"  
[591] "Sequoia National Park"  
[593] "Sequoia National Park"  
[595] "Sequoia National Park"  
[597] "Sequoia National Park"  
[599] "Sequoia National Park"  
[601] "Sequoia National Park"  
[603] "Sequoia National Park"  
[605] "Sequoia National Park"  
[607] "Sequoia National Park"  
[609] "Sequoia National Park"  
[611] "Sequoia National Park"  
[613] "Sequoia National Park"  
[615] "Sequoia National Park"

[617] "Sequoia National Park"  
[619] "Sequoia National Park"  
[621] "Sequoia National Park"  
[623] "Sequoia National Park"  
[625] "Sequoia National Park"  
[627] "Sequoia National Park"  
[629] "Sequoia National Park"  
[631] "Sequoia National Park"  
[633] "Sequoia National Park"  
[635] "Sequoia National Park"  
[637] "Sequoia National Park"  
[639] "Sequoia National Park"  
[641] "Sequoia National Park"  
[643] "Sequoia National Park"  
[645] "Sequoia National Park"  
[647] "Sequoia National Park"  
[649] "Sequoia National Park"  
[651] "Sequoia National Park"  
[653] "Sequoia National Park"  
[655] "Sequoia National Park"  
[657] "Sequoia National Park"  
[659] "Sequoia National Park"  
[661] "Sequoia National Park"  
[663] "Sequoia National Park"  
[665] "Sequoia National Park"  
[667] "Sequoia National Park"  
[669] "Sequoia National Park"  
[671] "Sequoia National Park"  
[673] "Sequoia National Park"  
[675] "Sequoia National Park"  
[677] "Sequoia National Park"  
[679] "Yosemite National Park"  
[681] "Yosemite National Park"  
[683] "Yosemite National Park"  
[685] "Yosemite National Park"  
[687] "Yosemite National Park"  
[689] "Yosemite National Park"  
[691] "Yosemite National Park"  
[693] "Yosemite National Park"  
[695] "Yosemite National Park"  
[697] "Yosemite National Park"  
[699] "Yosemite National Park"  
[701] "Yosemite National Park"

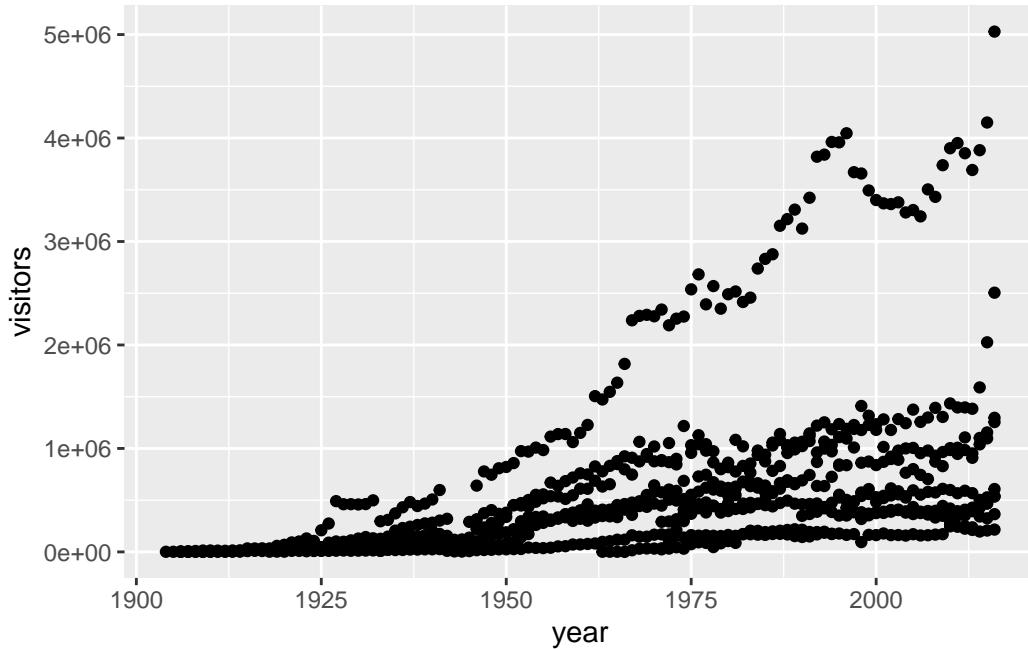


```
[789] "Yosemite National Park"
```

```
# Fetch the column but with the output run through the unique() function  
unique(ca_np$park_name)
```

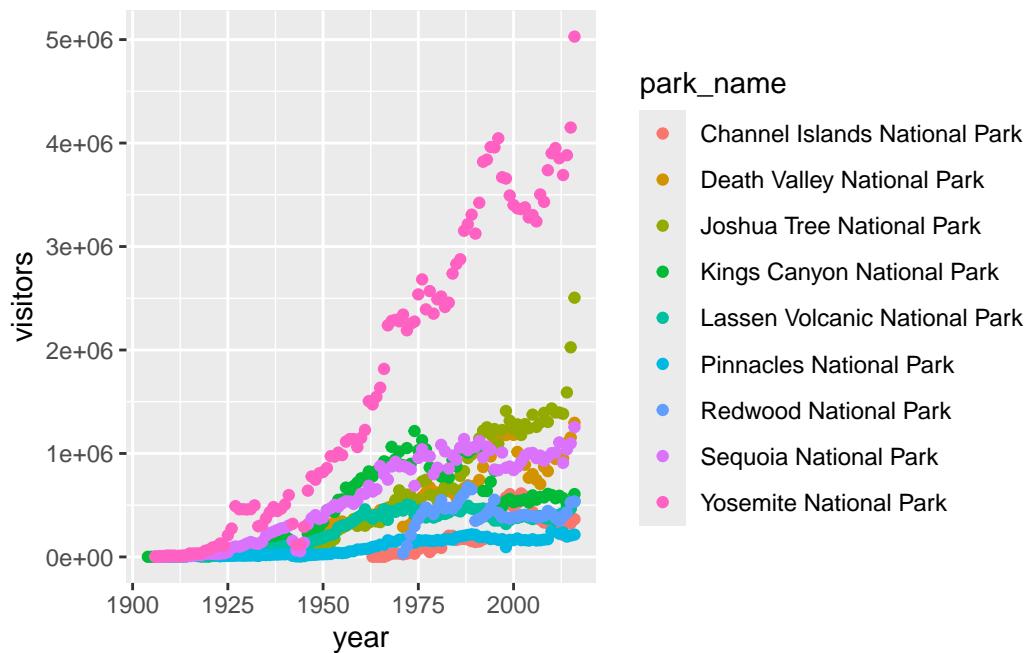
```
[1] "Channel Islands National Park" "Death Valley National Park"  
[3] "Joshua Tree National Park" "Kings Canyon National Park"  
[5] "Lassen Volcanic National Park" "Pinnacles National Park"  
[7] "Redwood National Park" "Sequoia National Park"  
[9] "Yosemite National Park"
```

```
# Scatterplot for all parks  
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +  
  geom_point()
```

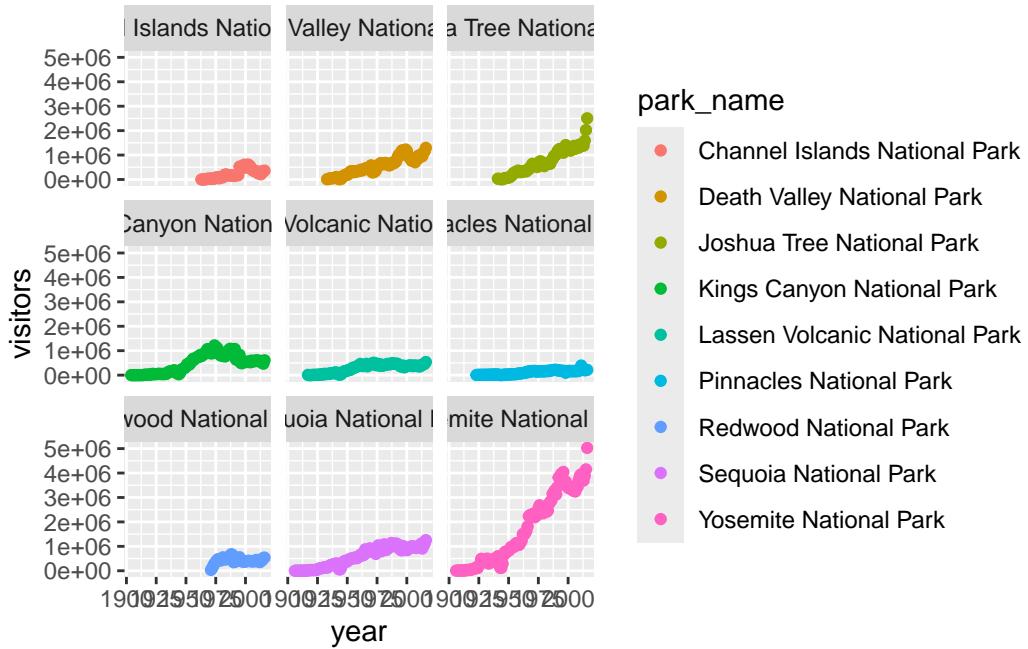


**Q5)**

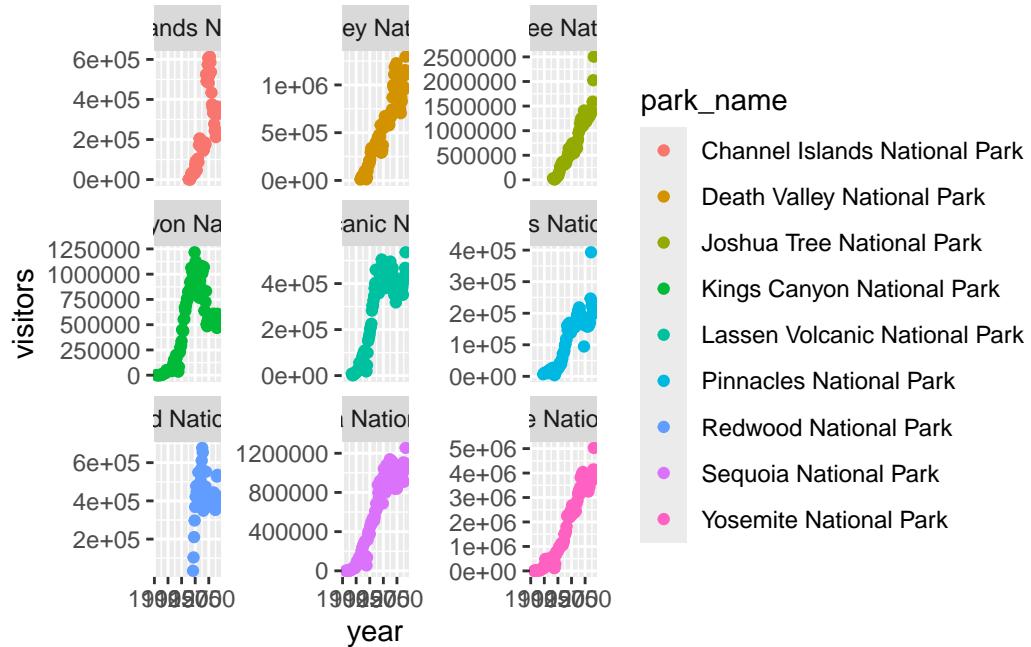
```
# Color coded the different parks  
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +  
  geom_point(aes(color = park_name))
```



```
# Added facet_wrap() to separate the different parks
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name)) +
  facet_wrap(park_name ~ .)
```



```
# changed the y-axis so each graph goes according to itself
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name)) +
  facet_wrap(park_name ~ ., scales = "free_y")
```



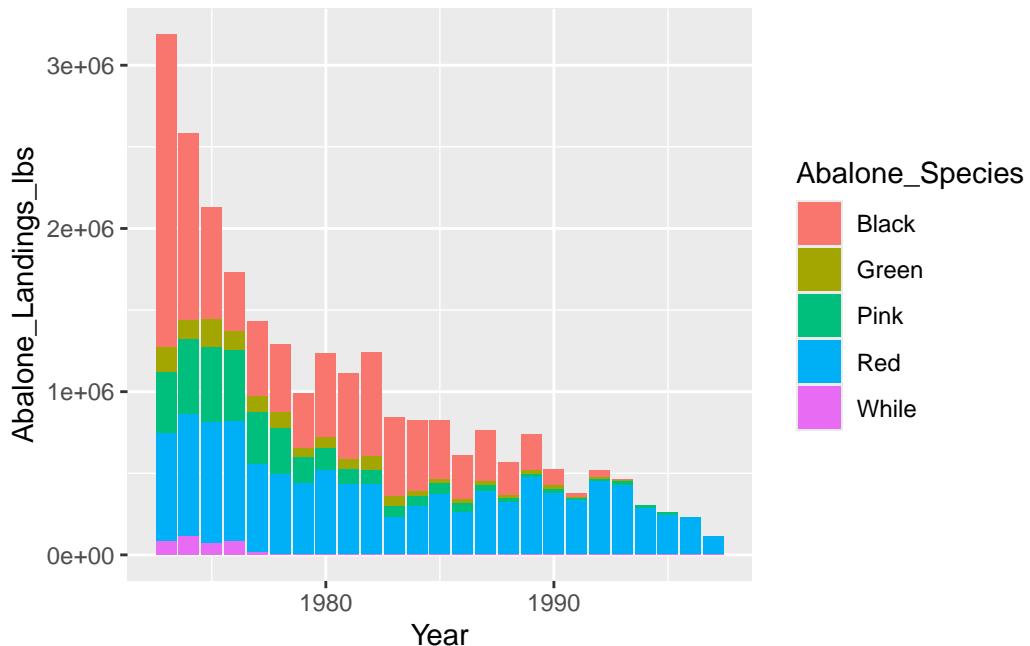
## Q6)

How does abalone landings differ based on species over a 25 year period?

## Q7)

```
abalone_landings <- read.csv("abalone_landings.csv")

abalone_landings <- read.csv("abalone_landings.csv")
ggplot(abalone_landings, aes(fill = Abalone_Species , x = Year, y = Abalone_Landings_lbs)) +
  geom_bar(position = "stack", stat = "identity")
```



- 1) I have learned from the graph that in general all of the abalone species surveyed are decreasing. However, some are decreasing at more dramatic rates than others.
- 2) How does abalone landing change based on location? This would require data on where the counts were conducted.