Activity 3: Data Visualization — Fundamentals of ggplot

Jake and Seleona

2026-01-12

install.packages("ggplot2")

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

```
here() starts at /Users/jacobmetzger/Documents/School/UCSC/Years/25to26/BIOE276 ds4eeb/Wk2/Da
ggplot-USE THIS ONE
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v lubridate 1.9.4     v tibble    3.3.0
v purrr     1.0.4     v tidyr     1.3.1

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

Read in the data and store as "ci_np" data object

```r
ci_np <- read.csv("ci_np.csv")
```

display the first 6 rows

```r
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
```

```r
colnames(ci_np)
```

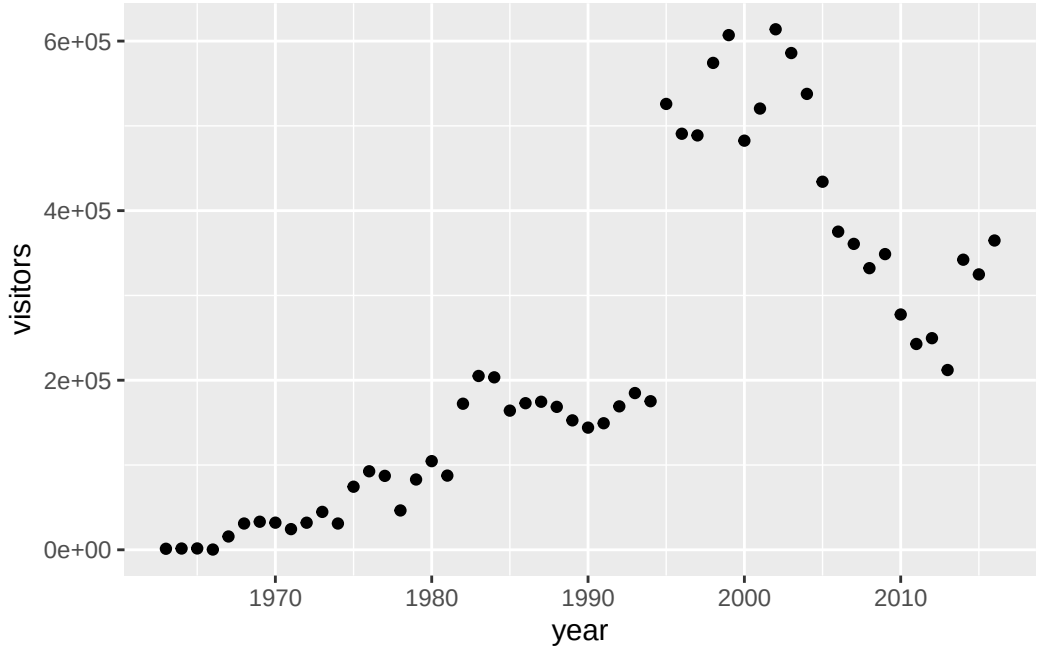
```
[1] "region"    "state"     "code"      "park_name" "type"      "visitors"
[7] "year"
```

```r
glimpse(ci_np)
```

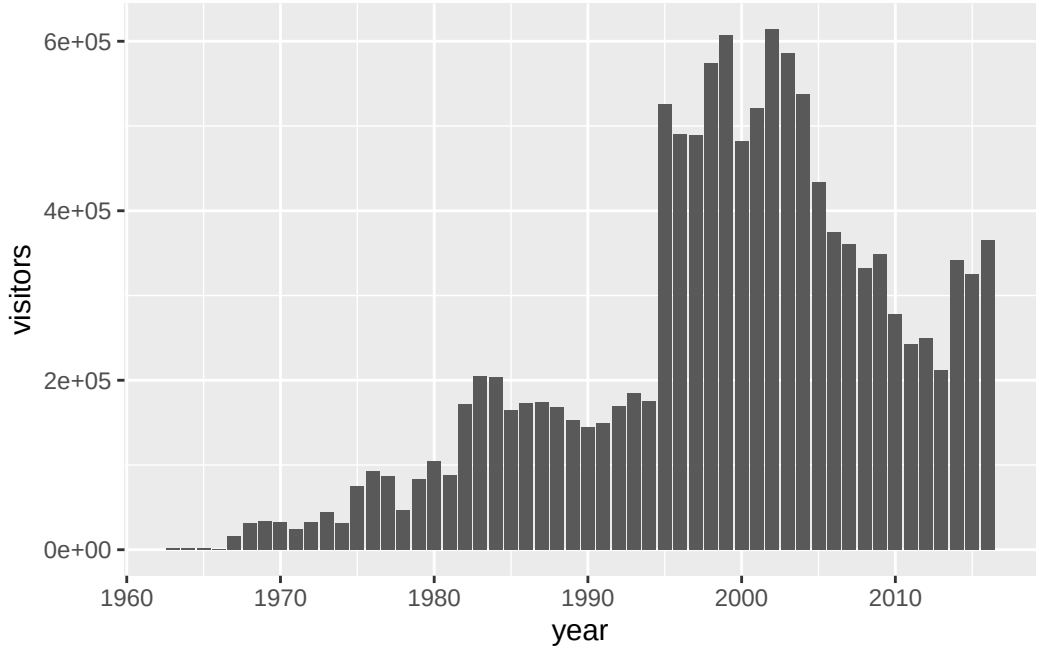
```
Rows: 54
Columns: 7
$ region    <chr> "PW", "PW", "PW", "PW", "PW", "PW", "PW", "PW", "PW", "PW", ~
$ state     <chr> "CA", "CA", "CA", "CA", "CA", "CA", "CA", "CA", "CA", "CA", ~
$ code      <chr> "CHIS", "CHIS", "CHIS", "CHIS", "CHIS", "CHIS", "CHIS", "CHI~
$ park_name <chr> "Channel Islands National Park", "Channel Islands National P~
$ type      <chr> "National Park", "National Park", "National Park", "National~
$ visitors  <int> 1200, 1500, 1600, 300, 15700, 31000, 33100, 32000, 24400, 31~
$ year      <int> 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, ~
```

```r
ggplot(data = ci_np)
```

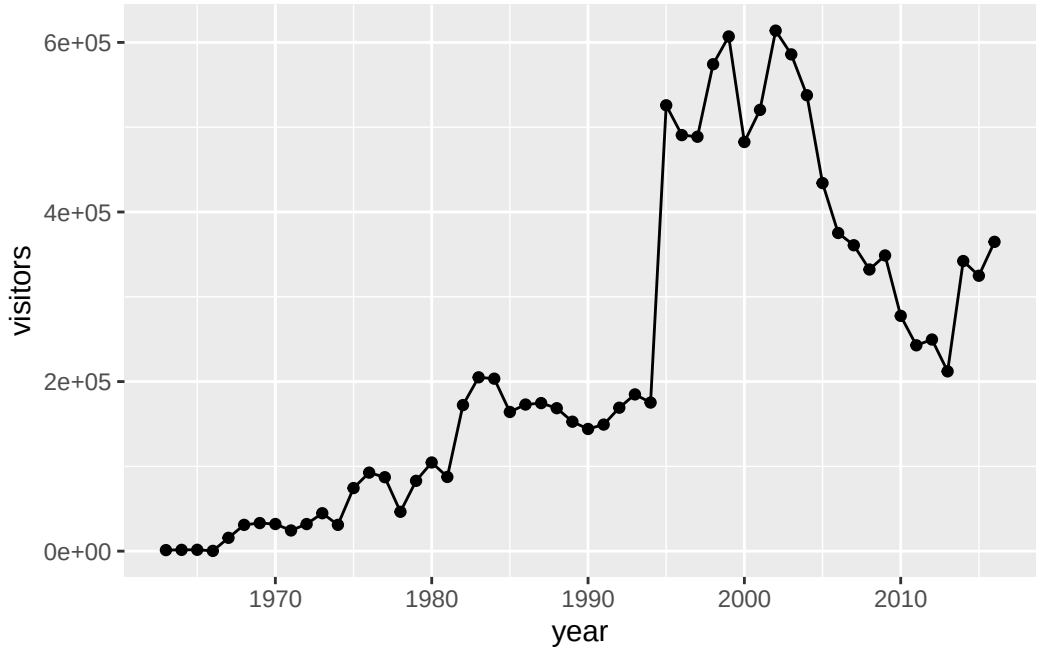
```
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point()
```



```
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_col()
```

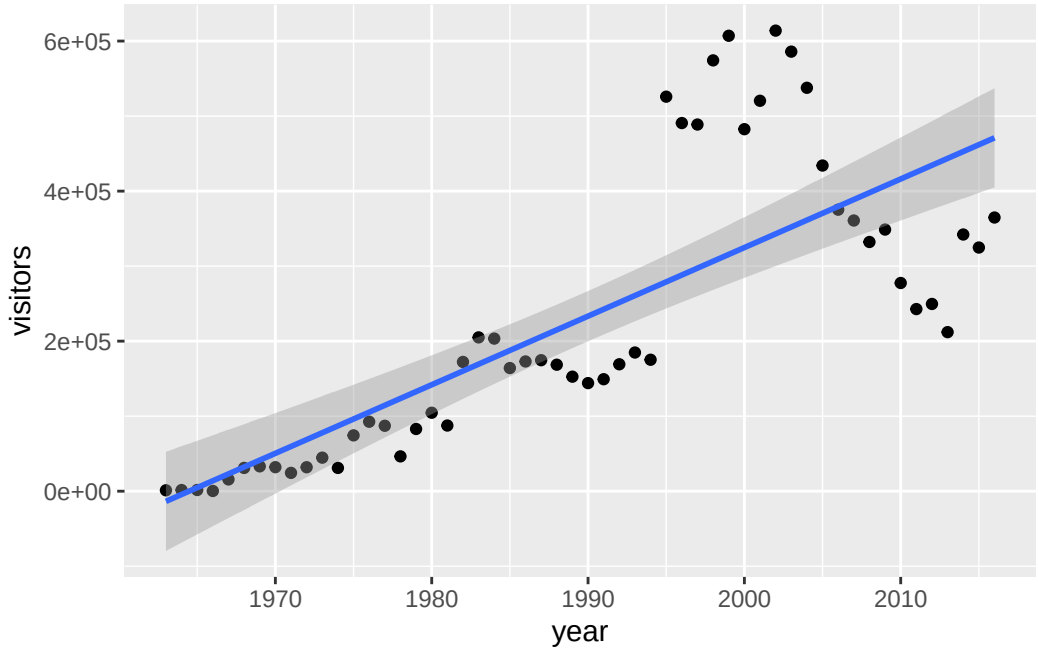



```
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point() +
  geom_line()
```



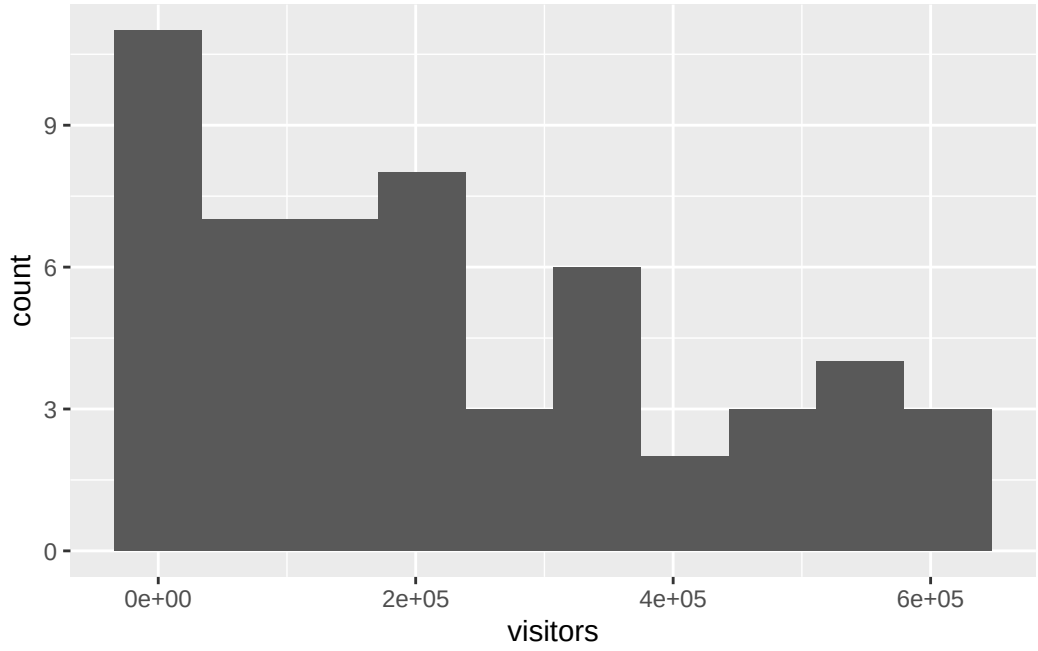

```
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point() +
  geom_smooth(method = 'lm' )
```

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

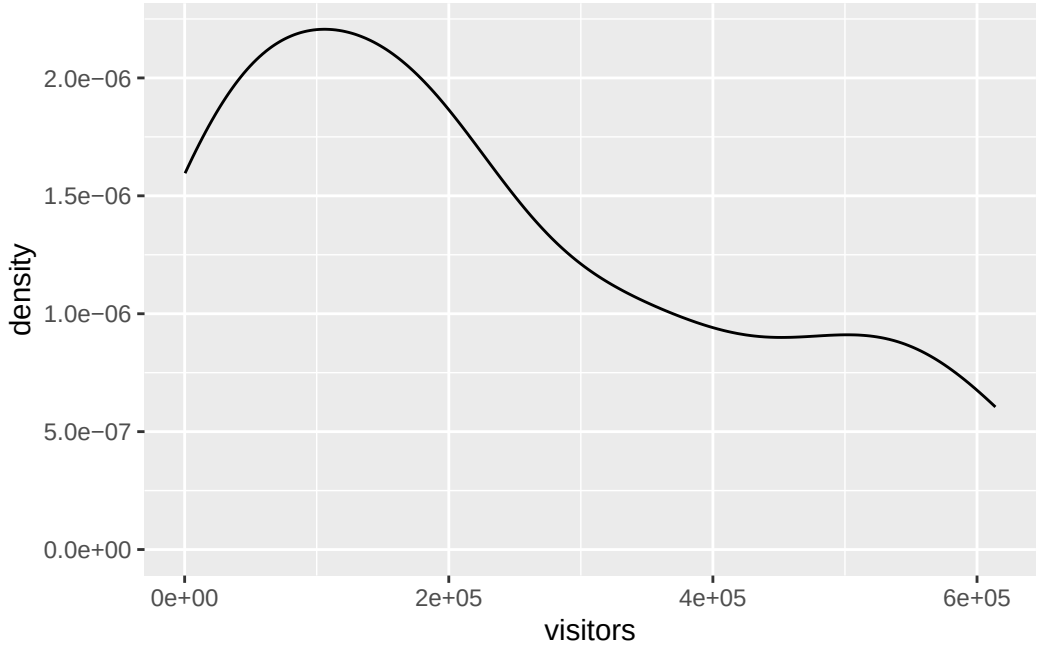


Plotting one variable —-

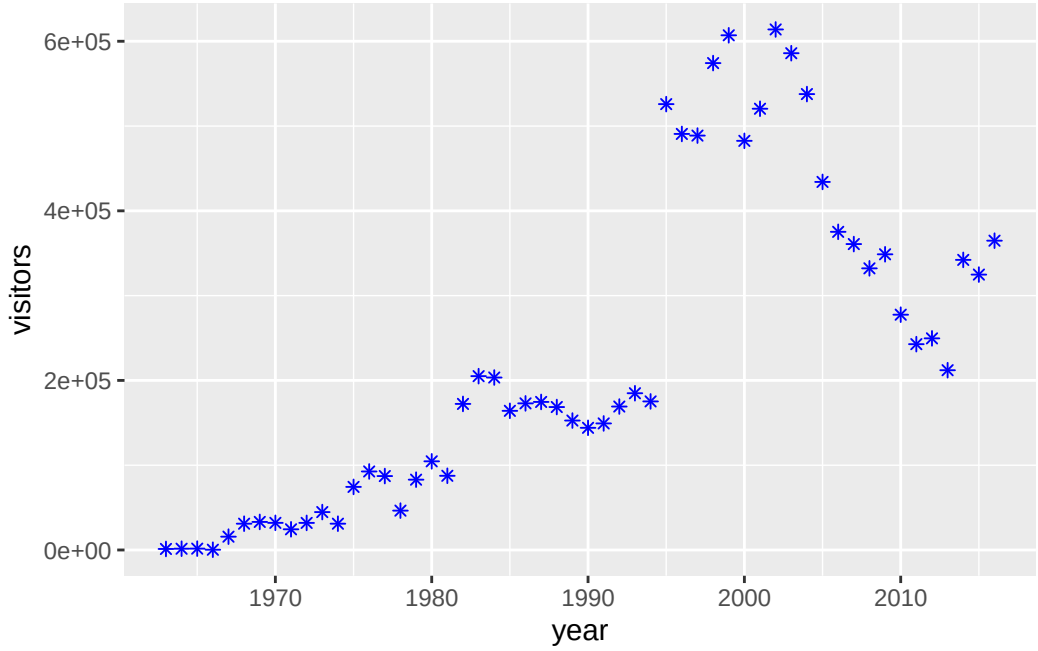
```
ggplot(data = ci_np, aes(x = visitors)) +
  geom_histogram(bins = 10)
```



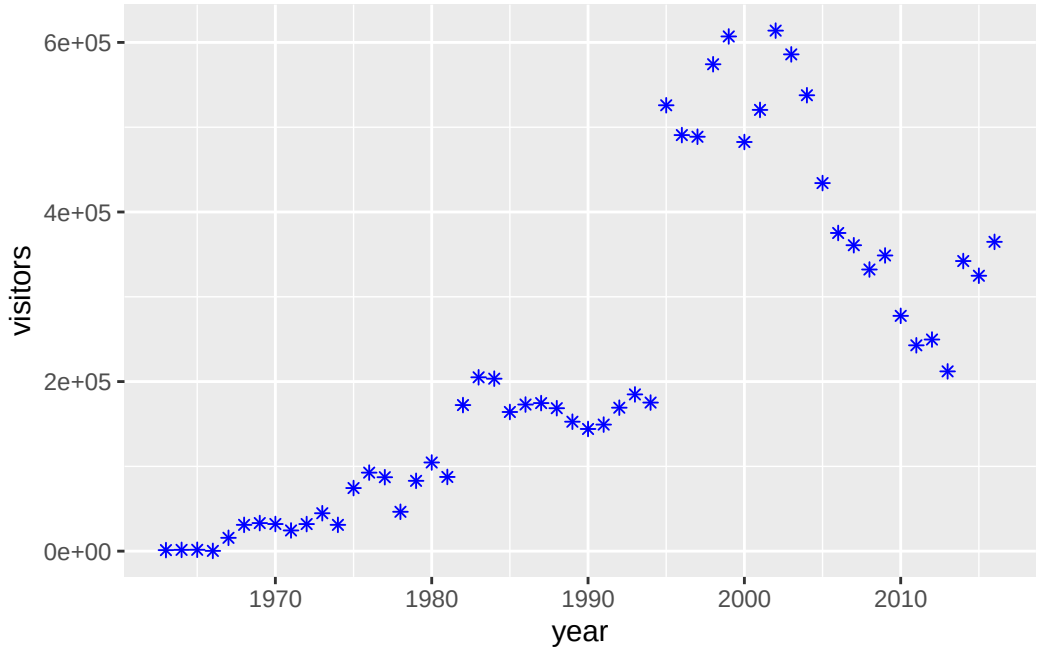

```
ggplot(data = ci_np, mapping = aes(x = visitors)) +
  geom_density()
```



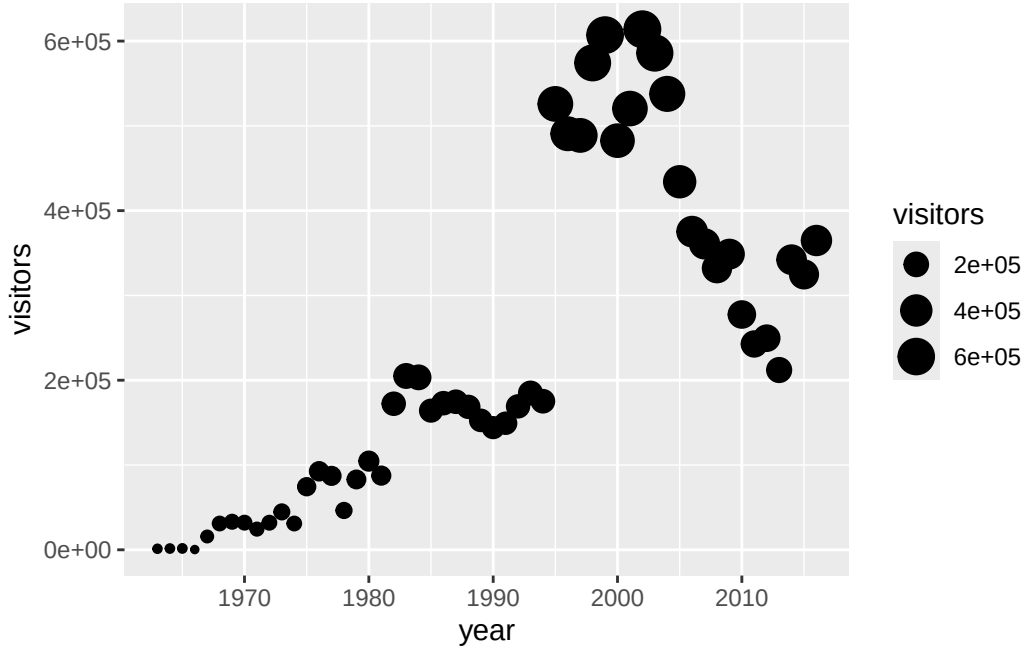
```
ggplot(data = ci_np, mapping = aes(x = year, y = visitors)) +
  geom_point(shape = 8, color = "blue")
```



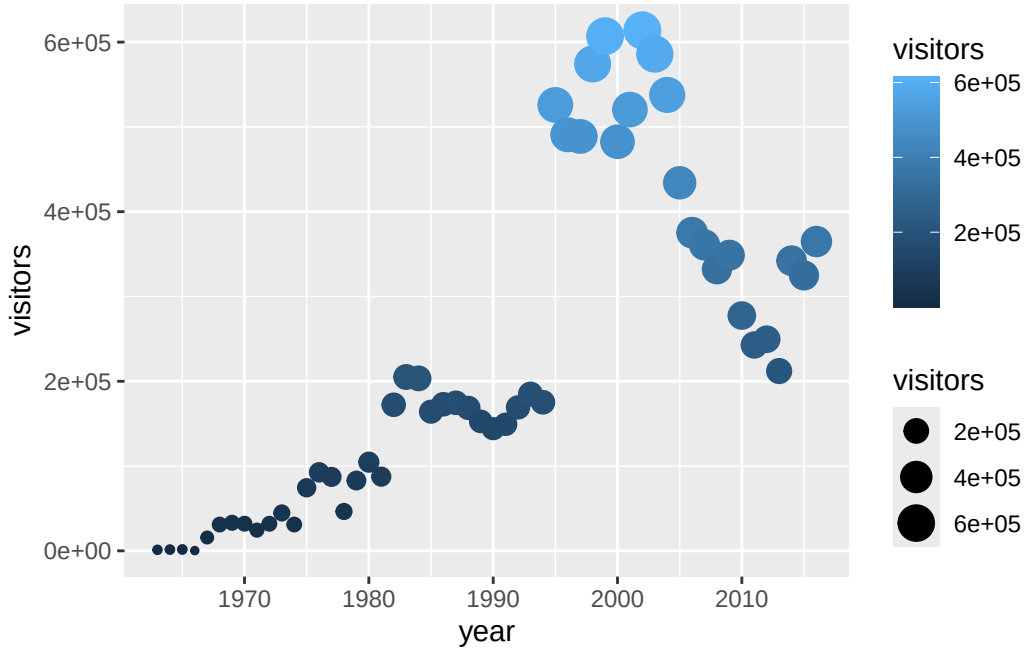

```
ggplot(ci_np, aes(x = year, y = visitors)) +
  geom_point(shape = 8, color = "blue")
```




```
ggplot(ci_np, aes(x = year, y = visitors)) +
  geom_point(aes(size = visitors))
```



```
ggplot(ci_np, aes(x = year, y = visitors)) +
  geom_point(aes(size = visitors,
                 color = visitors))
```

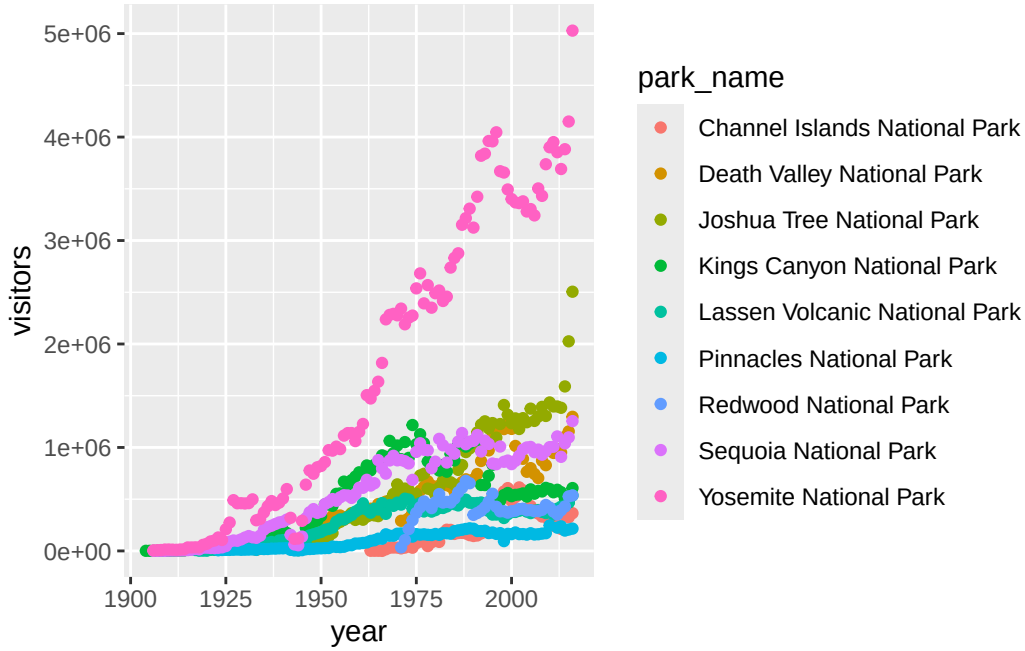


#Working with all park data

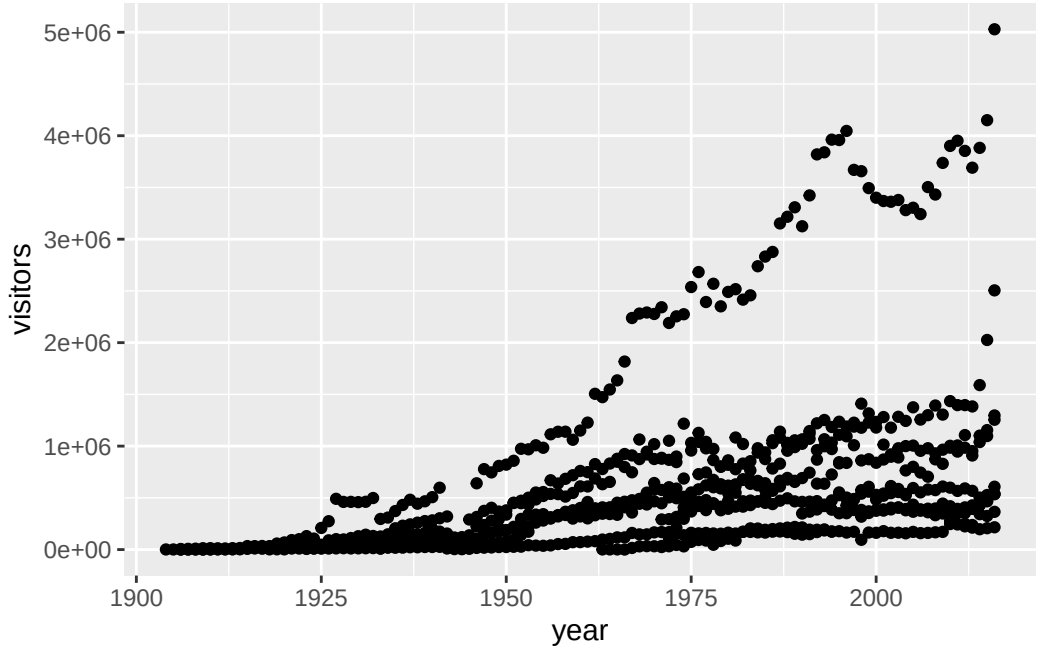
```
 ca_np <- read.csv("ca_np.csv")
unique(ca_np$park_name)
```

```
[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"
```

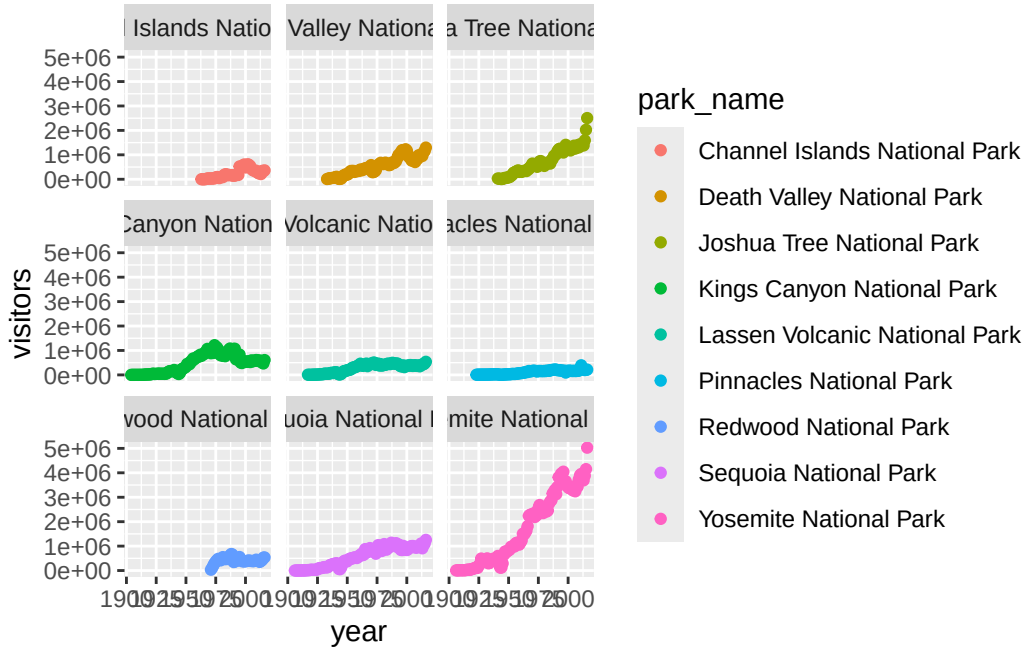
```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name))
```



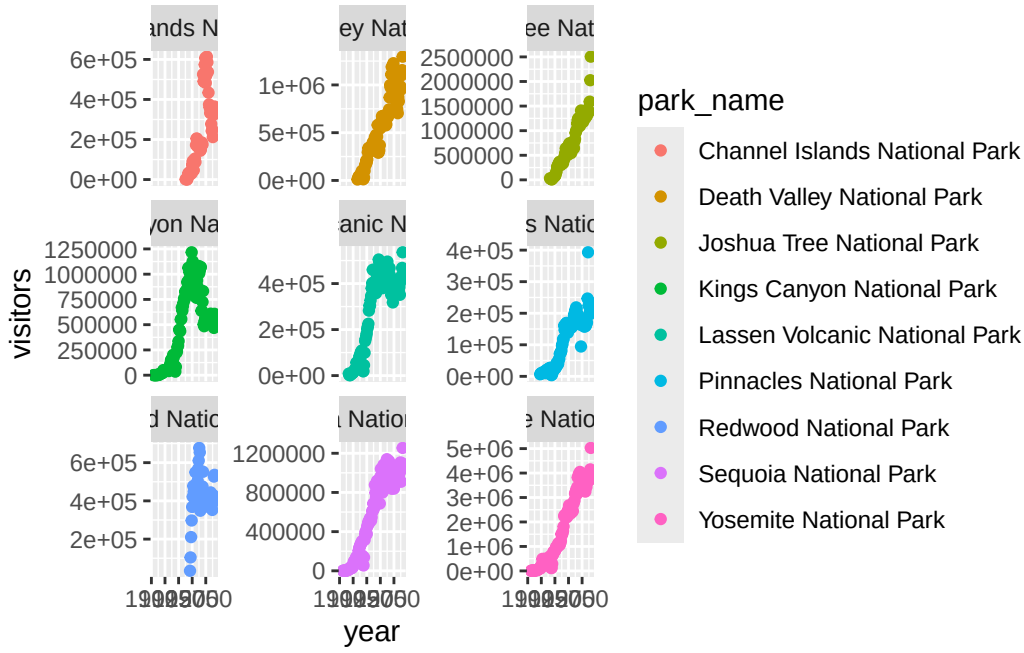

```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point()
```



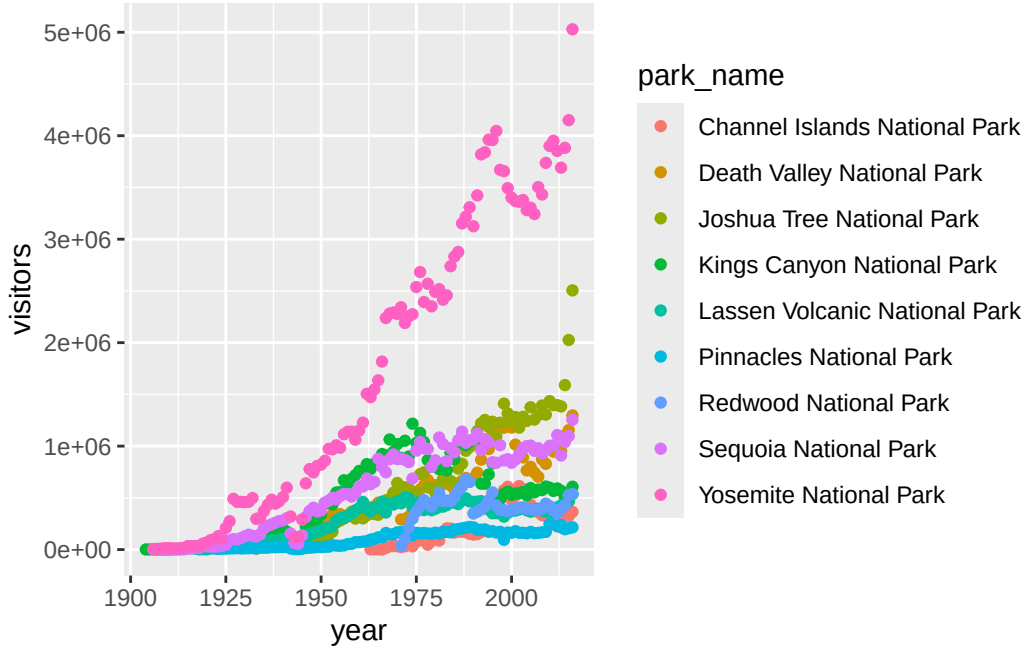
```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name)) +
  facet_wrap(park_name ~ .)
```



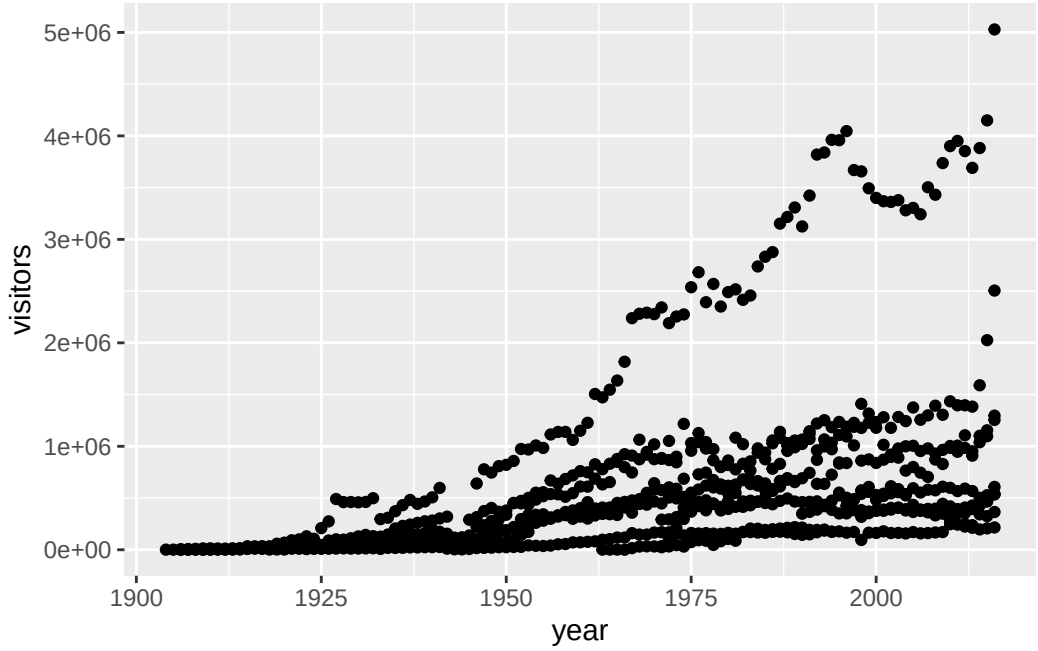

```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name)) +
  facet_wrap(park_name ~ ., scales = "free_y")
```



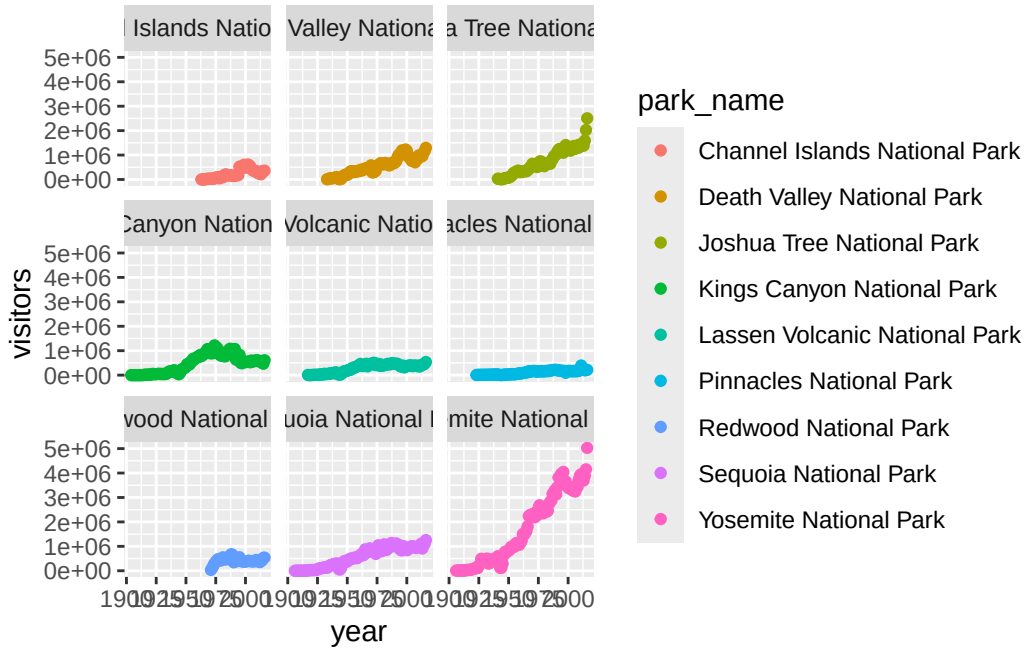

```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name))
```



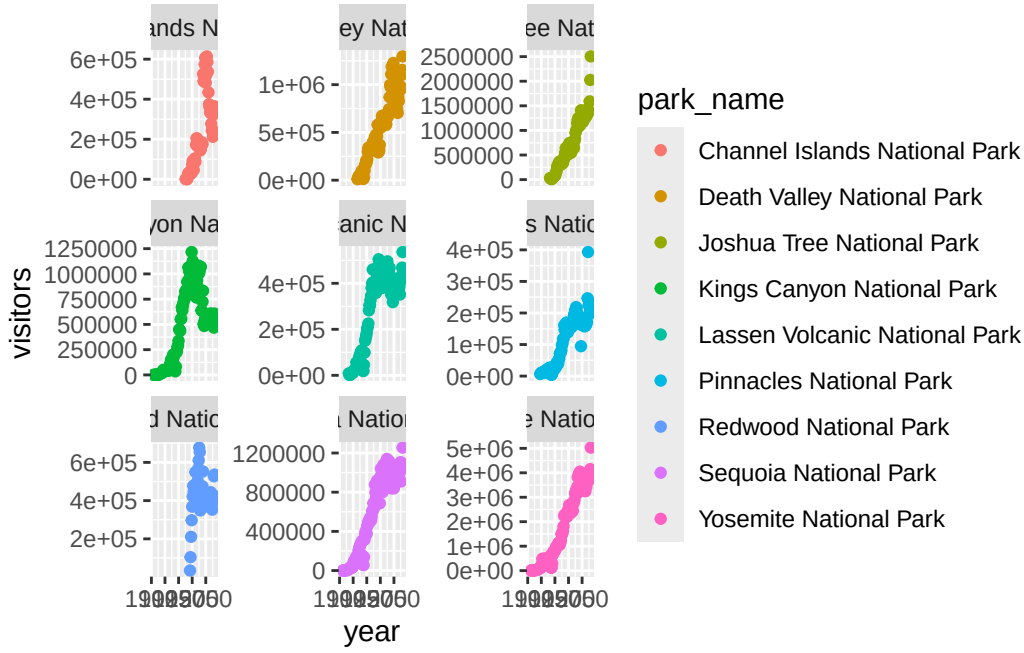
```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point()
```




```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name)) +
  facet_wrap(park_name ~ .)
```




```
ggplot(data = ca_np, mapping = aes(x = year, y = visitors)) +
  geom_point(aes(color = park_name)) +
  facet_wrap(park_name ~ ., scales = "free_y")
```



#Working with the abalone data

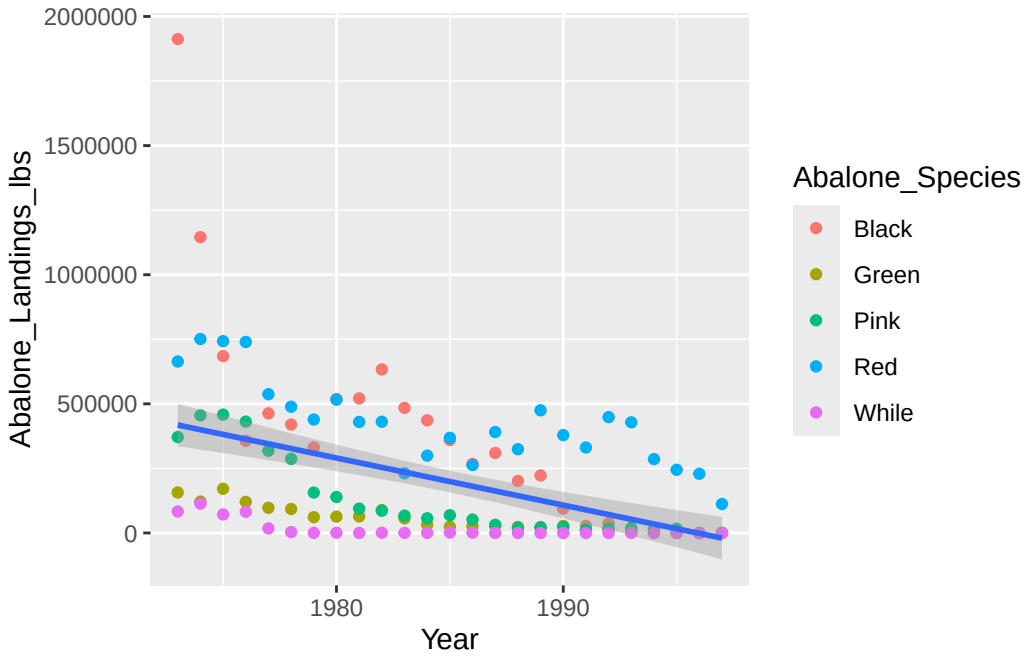
```
abalone <- read.csv("abalone_landings.csv")

glimpse(abalone)
```

```
Rows: 125
Columns: 3
$ Year                 <int> 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1~
$ Abalone_Species      <chr> "Black", "Black", "Black", "Black", "Black", "Bla~
$ Abalone_Landings_lbs <int> 1912519, 1145396, 684793, 356951, 463301, 420045,~
```

```
ggplot(data = abalone, mapping = aes(x = Year, y = Abalone_Landings_lbs)) +
  geom_point(aes(color = Abalone_Species))+
  geom_smooth(method = 'lm' )
```

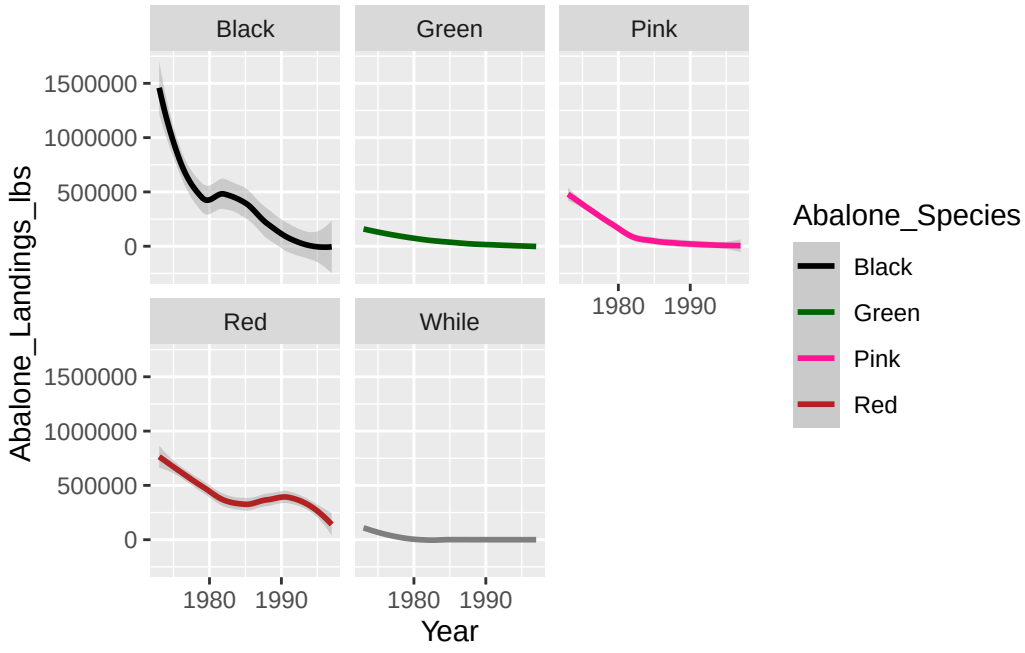
```
`geom_smooth()` using formula = 'y ~ x'
```



#Reseach question: All abalone species except Red abalone have 0 landings post 2001

```
ggplot(data = abalone, aes(x = Year, y = Abalone_Landings_lbs, color = Abalone_Species)) +
  geom_smooth() +
  facet_wrap(Abalone_Species ~ .) +
  scale_color_manual(values = c(
    "Red" = "firebrick",
    "Green" = "darkgreen",
    "Pink" = "deeppink",
    "Black" = "black",
    "White" = "white"
  ))
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggsave("abalone_landings_by_species.png",
       width = 8,
       height = 6,
       units = "in",
       dpi = 300)
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
unique(abalone$Abalone_Species)
```

```
[1] "Black" "Green" "Pink"  "Red"   "While"
```

What we learned from our graph:

All species declined precipitously over the course of the dataset. But, only red abalone had any harvest (>0lbs) at the endo of the data series.

Follow up research question:

How did juvenile abalone populations fare during this period. We would need size data, and potentially non-fisheries data to answer this question.