Lab 1

MGW

1/9/2020

Chunk 1

- Control Alt I opens a new R chunk.
- Here tells you to look in the following folder to find the file

```
us_landings<-read.csv(here("data","noaa_fisheries.csv"))
```

Chunk 2

- Now, make a tidy version of the data frame
- Janitor is cleaning up the dataframe
  - clean_names made everything lowercase
  - mutate: add an additional column that means something new
    - ..unless you want to change the existing variable!
    - put the existing name of the variable after mutate (in this case, state)
    - apply it to the existing variable, state (after str_to_lower)
    - this puts everything in lowercase!

```
landings_tidy <- us_landings%>%
  janitor::clean_names()%>%
  mutate(state=str_to_lower(state),
         afs_name=str_to_lower(afs_name))
```

Chunk 3

- In the readr package, there's a bunch of options for parsing things, such as parse number. If you only want to get the numeric version of an amount of money (such as only wanting 50 from $50), use parse number in readr.

```
landings_tidy<-landings_tidy%>%
  mutate(dollars_num=parse_number(as.character(dollars_usd)))
```

Chunk 4

- Remove the word "aggregate" from all rows in the column afs_name
- `str_detect(salmon_landings$afs_clean,pattern="salmon")` tells you if each row has the word 'salmon' in it or not for that variable
- Now, let's filter out every row with the word salmon into two groups using the command `separate`
- This created a new dataset called salmon_landings with only 1500 obs

```
salmon_landings <- landings_tidy%>%
  mutate(afs_clean=str_remove(afs_name, pattern="aggregate"))%>%
```

```
  filter(str_detect(afs_clean, pattern="salmon"))%>%
  separate(afs_clean, into=c("group","species"),sep=",")
```

Chunk 5

- Find some grouped summary data:
- Find annual total US landings and dollar value (summing across all states) for each type of salmon using `group_by` and `summary`
- Reduces to <500 rows because there's like 60 years and a certain number of species per year

```
salmon_summary<-salmon_landings%>%
  group_by(year, species) %>%
  summarize(
    tot_landings = sum(landings_pounds),
    tot_value = sum(dollars_num)
  )
```

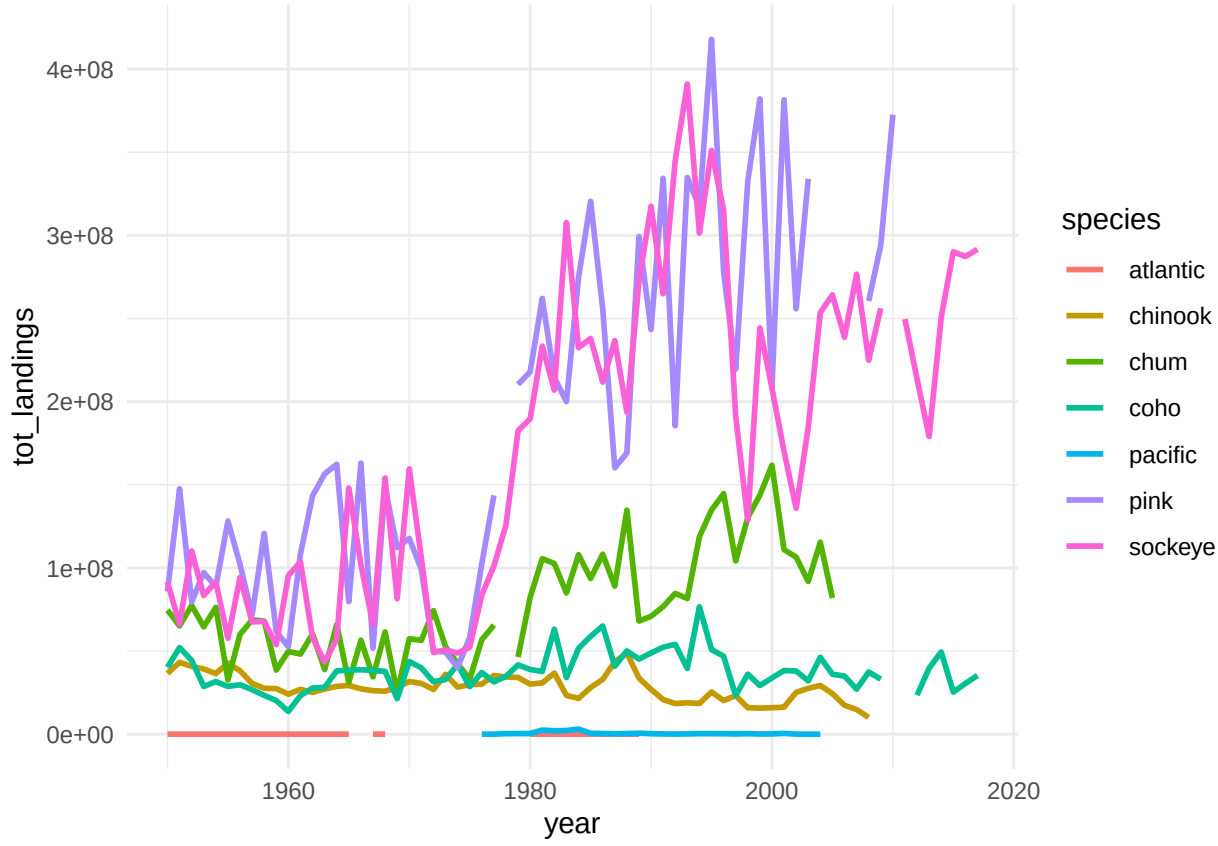
Chunk 6

- Now, let's plot
- group gives different lines for each group (could also do color = species) in either the first aesthetic or the second aesthetic

```
salmon_landings_graph <- ggplot(data=salmon_summary,
                                aes(x=year,y=tot_landings))+
  geom_line(aes(color=species),size=1)+
  theme_minimal()

salmon_landings_graph
```

```
## Warning: Removed 22 rows containing missing values (geom_path).
```



Chunk 7

- Export graph with certain dimensions for journal

```
ggsave(plot=salmon_landings_graph,
       here("figures","mygraph.png"),
       height=5,width=8)
```

```
## Warning: Removed 22 rows containing missing values (geom_path).
```

Chunk 8

- Make a nice kable table:

```
salmon_first_5 <- salmon_summary %>%
  head(5)

kable(salmon_first_5)%>%
  kable_styling(bootstrap_options="striped",
                full_width=FALSE)
```

year	species	tot_landings	tot_value
1950	atlantic	800	362
1950	chinook	36559160	8328433
1950	chum	74610432	5920158
1950	coho	40224632	7151008
1950	pink	85773432	6773180