

Assignment 6 Part 2

Gyaban Essilfie-Bondzie

2023-03-27

#Intermezzo 9.5

```
library(tidyverse)

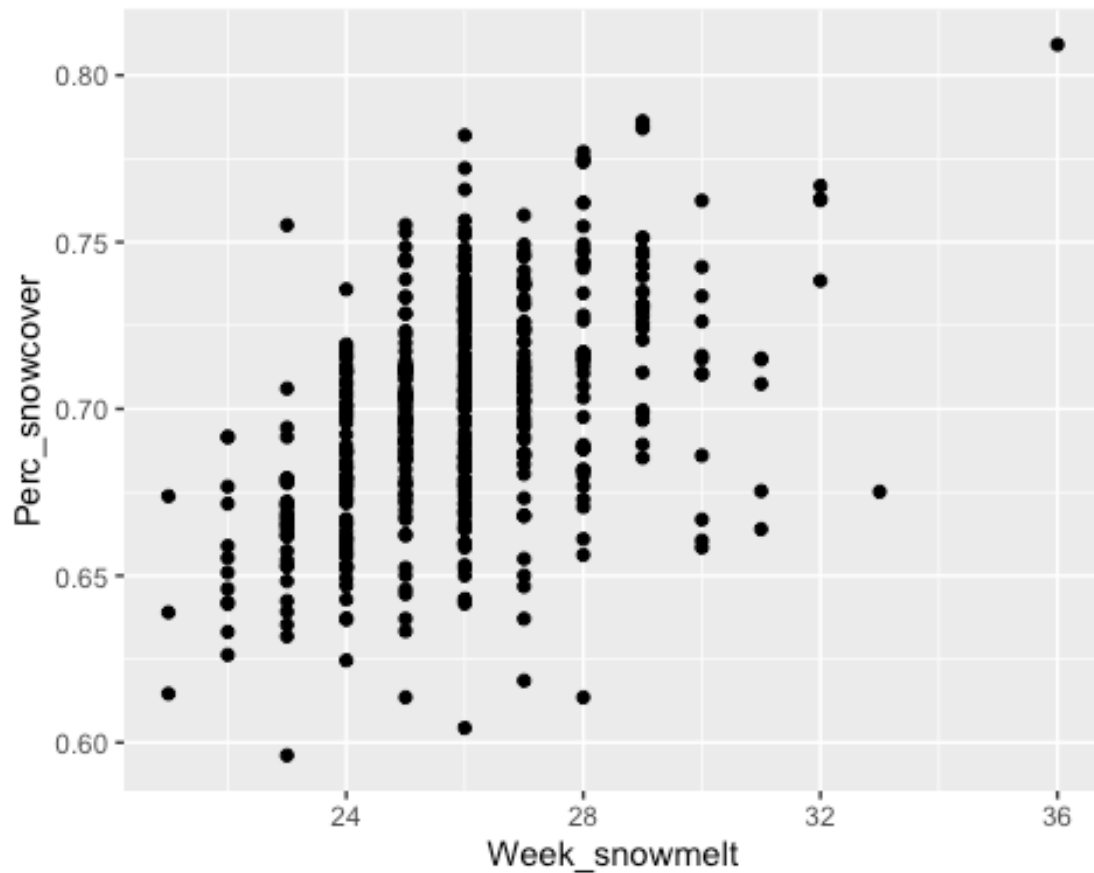
## — Attaching packages — tidyverse
1.3.2 —
## ✓ ggplot2 3.4.1    ✓ purrr 1.0.1
## ✓ tibble 3.1.8     ✓ dplyr 1.1.0
## ✓ tidyr 1.3.0      ✓ stringr 1.5.0
## ✓ readr 2.1.4      ✓ forcats 1.0.0
## — Conflicts —
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag() masks stats::lag()

#Load file
snow <- read_tsv("data_wrangling/data/FauchaldEtAl2017/snow.csv")

## Rows: 495 Columns: 4
## — Column specification
##
## Delimiter: "\t"
## chr (1): Herd
## dbl (3): Year, Perc_snowcover, Week_snowmelt
##
## ⓘ Use `spec()` to retrieve the full column specification for this data.
## ⓘ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

#graph week_snowmelt vs perc_snowcover
snow %>%
  ggplot(aes(x=Week_snowmelt, y=Perc_snowcover)) + geom_point()

## Warning: Removed 26 rows containing missing values (`geom_point()`).
```



```
#Plot the avg Perc_snowcover vs avgWeeksnowmelt grouped by year
snow %>%
  summarise(
    avgPerc_snowcover = mean(Perc_snowcover), avgWeekSnowMelt =
    mean(Week_snowmelt)) %>%
    ggplot(aes(x = avgWeekSnowMelt, y = avgPerc_snowcover)) + geom_point()
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```

avgPerc_snowcover

avgWeekSnowMelt

#Boxplot for the population size of each herd

```
popsize <- read_tsv("data_wrangling/data/FauchaldEtAl2017/pop_size.csv")
```

```
## Rows: 114 Columns: 3
```

```
## — Column specification
```

```
## Delimiter: "\t"
```

```
## chr (1): Herd
```

```
## dbl (2): Year, Pop_Size
```

```
##
```

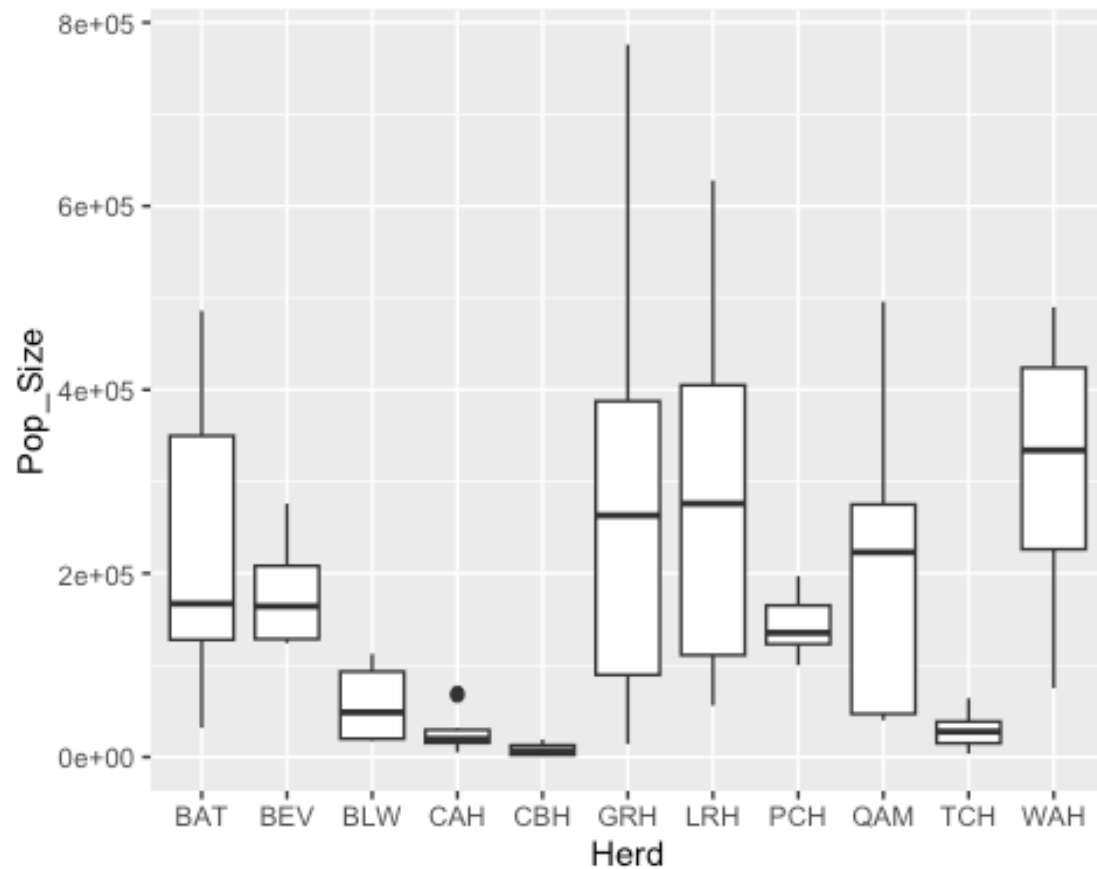
```
## ⓘ Use `spec()` to retrieve the full column specification for this data.
```

```
## ⓘ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
popsize %>%
```

```
  group_by(Herd) %>%
```

```
  ggplot(aes( x = Herd, y = Pop_Size)) + geom_boxplot()
```



#Boxplot for the population size of each herd across the years 2008-2014

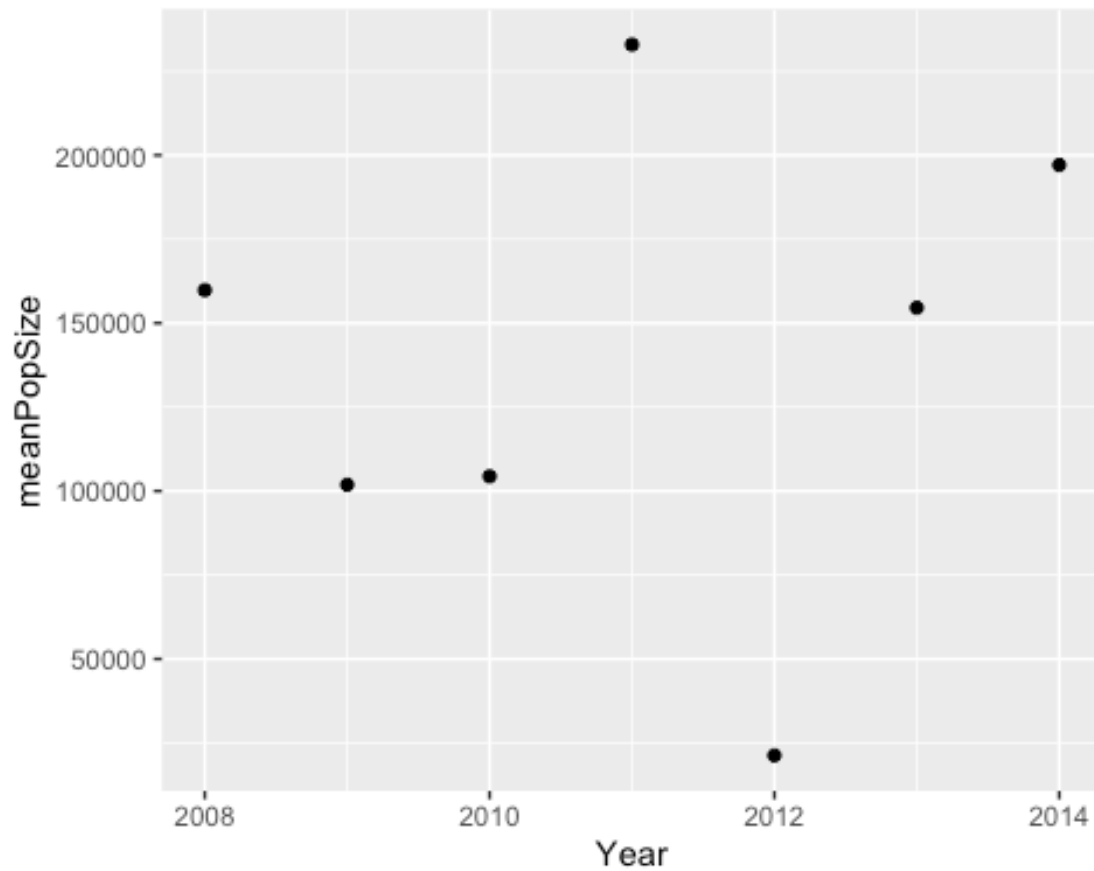
popsiz %>%

filter(Year >= 2008, Year <= 2014) %>%

group_by(Year) %>%

summarise(meanPopSize= mean(Pop_Size), SD = sd(Pop_Size)) %>%

ggplot(aes(x =Year, y = meanPopSize)) + geom_point()



#Intermezzo 9.6

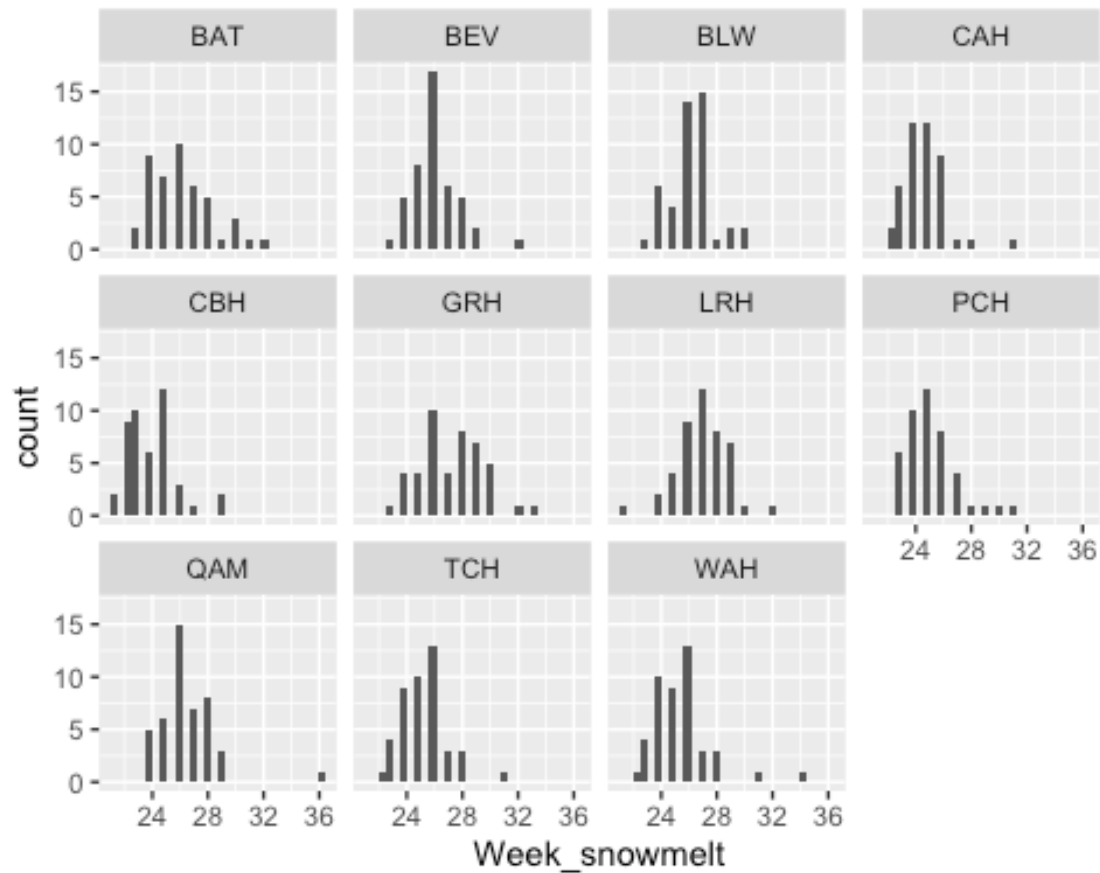
#Graph each herd a histogram of weekly snowmelt

snow %>%

```
ggplot(aes(x = Week_snowmelt)) + geom_histogram() +  
  facet_wrap(~Herd)
```

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

Warning: Removed 3 rows containing non-finite values (`stat_bin()`).



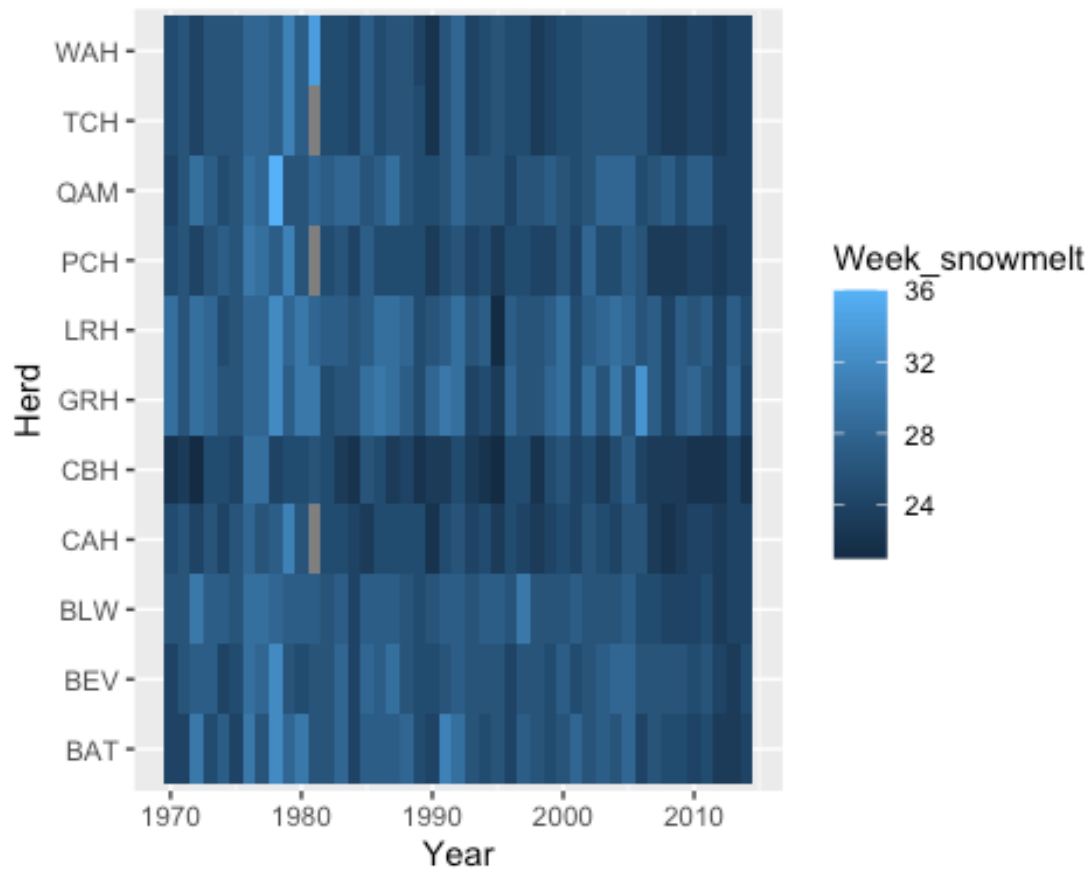
```
#Explore geom_tile
```

```
?geom_tile
```

```
#Heat map with x as year and y as Herd
```

```
snow %>%
```

```
  ggplot(aes( x =Year, y = Herd, fill = Week_snowmelt)) + geom_tile()
```



#graph week_snowmelt vs perc_snowcover and add a smooth line

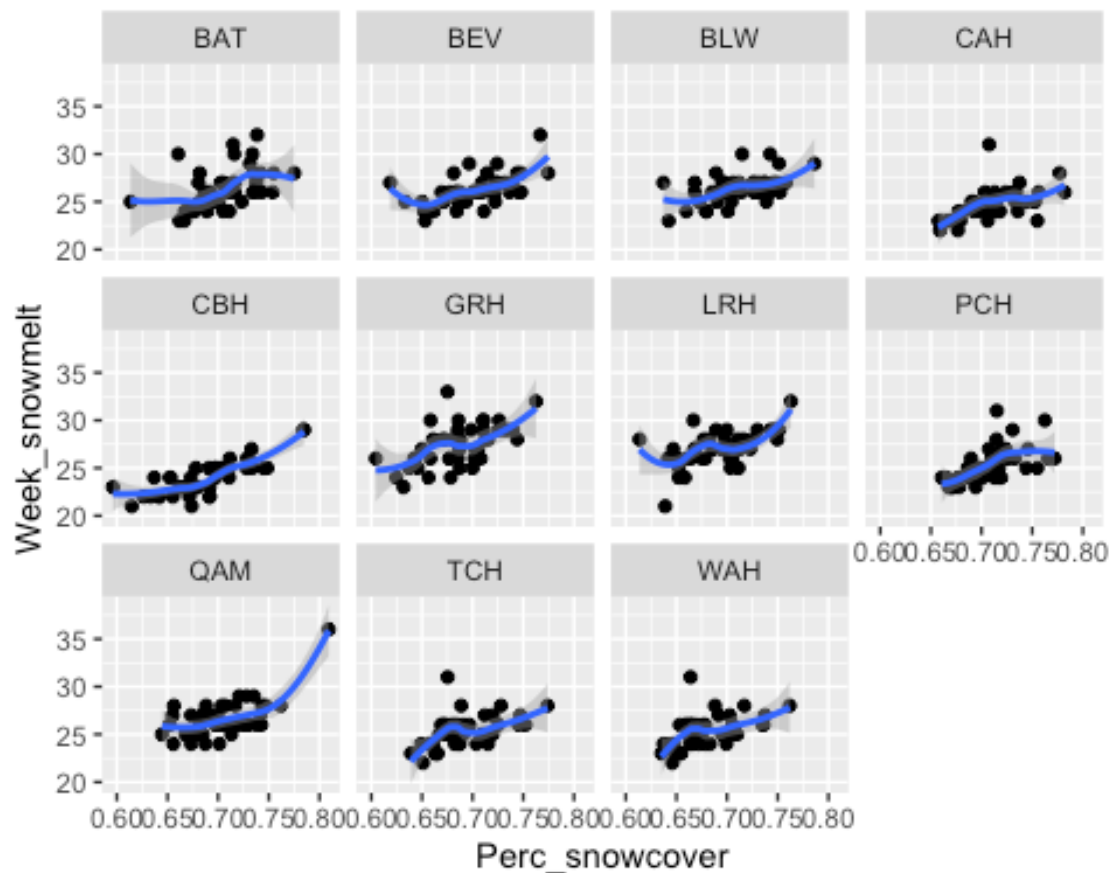
snow %>%

```
ggplot(aes(y=Week_snowmelt, x=Perc_snowcover)) + geom_point() +  
geom_smooth() +  
facet_wrap(~Herd)
```

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

Warning: Removed 26 rows containing non-finite values (`stat_smooth()`).

Warning: Removed 26 rows containing missing values (`geom_point()`).



```
#Test
```

```
#Load data
```

```
data(iris)
```

```
#Perform ttest between the 2 species
```

```
iris %>%
```

```
  filter(Species != 'virginica') %>%
```

```
  t.test(data = ., Sepal.Length ~ Species)
```

```
##
```

```
##  Welch Two Sample t-test
```

```
##
```

```
## data: Sepal.Length by Species
```

```
## t = -10.521, df = 86.538, p-value < 2.2e-16
```

```
## alternative hypothesis: true difference in means between group setosa and  
## group versicolor is not equal to 0
```

```
## 95 percent confidence interval:
```

```
##  -1.1057074 -0.7542926
```

```
## sample estimates:
```

```
##      mean in group setosa mean in group versicolor
```

```
##              5.006
```

```
              5.936
```


#Interpret

#The null hypothesis is the hypothesis that there is no significant difference between the Sepal Length of the species setosa & versicolor
#The P value measures how likely it is that any observed difference between groups is due to chance.

Yes, the null hypothesis is rejected when p-value is Less than 5%

#Visualize sepal width of the two species using a boxplot

`iris %>%`

`filter(Species != 'versicolor') %>%`

`ggplot(aes(x=Species, y=Sepal.Width)) + geom_boxplot()`

