

# HW2 - Birds Analysis

John C. Manyiel

9/15/2021

Load packages

```
library(rprojroot)
#root<-has_file(".ROS-Examples-root")$make_fix_file()
library(rstanarm)
library(arm)
library(ggplot2)
library(bayesplot)
library(tidyverse)
library(stringr)
library(pdftools)
library(stringr)
library(xlsx)
theme_set(bayesplot::theme_default(base_family = "sans"))
```

Load Data

```
df <- read.csv("Z:/Development/DatumAnalyticConsulting/DACSS601Fall21/_data/birds.csv")
```

head(df)

	ï..Domain.Code	Domain	Area.Code	Area	Element.Code	Element
1	QA Live Animals	2 Afghanistan	5112	Stocks		
2	QA Live Animals	2 Afghanistan	5112	Stocks		
3	QA Live Animals	2 Afghanistan	5112	Stocks		
4	QA Live Animals	2 Afghanistan	5112	Stocks		
5	QA Live Animals	2 Afghanistan	5112	Stocks		
6	QA Live Animals	2 Afghanistan	5112	Stocks		

  

	Item.Code	Item	Year.Code	Year	Unit	Value	Flag	Flag.Description
1	1057 Chickens	1961 1961	1000	Head	4700	F	FAO estimate	
2	1057 Chickens	1962 1962	1000	Head	4900	F	FAO estimate	
3	1057 Chickens	1963 1963	1000	Head	5000	F	FAO estimate	
4	1057 Chickens	1964 1964	1000	Head	5300	F	FAO estimate	
5	1057 Chickens	1965 1965	1000	Head	5500	F	FAO estimate	
6	1057 Chickens	1966 1966	1000	Head	5800	F	FAO estimate	

Learn & Explore the data set

```
df_sum <- summary(df)
```

df\_sum

ï..Domain.Code	Domain	Area.Code	Area
----------------	--------	-----------	------

```

Length:30977      Length:30977      Min.   :    1      Length:30977
Class :character   Class :character   1st Qu.:   79      Class :character
Mode  :character   Mode  :character   Median :  156      Mode  :character
                        Mean   : 1202
                        3rd Qu.:  231
                        Max.   : 5504

```

```

Element.Code      Element              Item.Code      Item
Min.   :5112      Length:30977      Min.   :1057      Length:30977
1st Qu.:5112      Class :character   1st Qu.:1057      Class :character
Median :5112      Mode  :character   Median :1068      Mode  :character
Mean   :5112                        Mean   :1066
3rd Qu.:5112                        3rd Qu.:1072
Max.   :5112                        Max.   :1083

```

```

Year.Code      Year      Unit      Value
Min.   :1961      Min.   :1961      Length:30977      Min.   :    0
1st Qu.:1976      1st Qu.:1976      Class :character   1st Qu.:   171
Median :1992      Median :1992      Mode  :character   Median :   1800
Mean   :1991      Mean   :1991                        Mean   :  99411
3rd Qu.:2005      3rd Qu.:2005                        3rd Qu.: 15404
Max.   :2018      Max.   :2018                        Max.   :23707134
                        NA's   :1036

```

```

Flag      Flag.Description
Length:30977      Length:30977
Class :character   Class :character
Mode  :character   Mode  :character

```

```

dfdetails<- str(df)
'data.frame':  30977 obs. of  14 variables:
 $ i..Domain.Code : chr  "QA" "QA" "QA" "QA" ...
 $ Domain         : chr  "Live Animals" "Live Animals" "Live Animals" "Live Animals" ...
 $ Area.Code      : int   2 2 2 2 2 2 2 2 2 2 ...
 $ Area          : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ Element.Code   : int   5112 5112 5112 5112 5112 5112 5112 5112 5112 ...
 $ Element        : chr  "Stocks" "Stocks" "Stocks" "Stocks" ...
 $ Item.Code      : int   1057 1057 1057 1057 1057 1057 1057 1057 1057 ...
 $ Item          : chr  "Chickens" "Chickens" "Chickens" "Chickens" ...
 $ Year.Code      : int   1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 ...
 $ Year          : int   1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 ...
 $ Unit          : chr  "1000 Head" "1000 Head" "1000 Head" "1000 Head" ...
 $ Value         : int   4700 4900 5000 5300 5500 5800 6600 6290 6300 6000 ...
 $ Flag          : chr  "F" "F" "F" "F" ...
 $ Flag.Description: chr  "FAO estimate" "FAO estimate" "FAO estimate" "FAO estimate" ...
dfdetails
NULL

```

Select random from Bird Data

```
df_smaple <- sample_frac(df, 0.1)
head(df_smaple[1:5])
```

	i..Domain.Code	Domain	Area.Code	Area
1		QA Live Animals	105	Israel
2		QA Live Animals	27	Bulgaria
3		QA Live Animals	215	United Republic of Tanzania
4		QA Live Animals	63	Estonia
5		QA Live Animals	5204	Central America
6		QA Live Animals	162	Norway

  

	Element.Code
1	5112
2	5112
3	5112
4	5112
5	5112
6	5112

Data Transformation & Wrangling: discovered that we would need to transform the data before.

```
df1 <- as_tibble(df) %>%
  group_by(Area) %>%

  filter(
    Area == "Northern America" |
    Area == "United States of America" |
    Area == "Canada" |
    Area == "Mexico" |
    Area == "American Samoa" |
    Area == "United States Virgin Islands"
  ) %>% summarise(sum(Value))
```

```
df1
# A tibble: 6 x 2
  Area                                `sum(Value)`
<chr>                                <int>
1 American Samoa                     2402
2 Canada                             7314967
3 Mexico                             17312588
4 Northern America                   100752122
5 United States of America           93432842
6 United States Virgin Islands        2387
```

Notice that North America value is combination of all the states. I would need to transform the data further.