STA5092Z: Assigment 1

Yevashan Perumal

05/04/2021

Summary of data

Data was collected for each breeding season from September 2005 - February 2019 for the Southern Pied Babbler. The data consists of 6 csv files; 3 files containing data and 3 files containing metadata describing the fields in each data set. Below we import, wrangle and explore some of the data.

Part 1 - Data Wrangling

A preview of the dataset named Data.Bourne_NestlingsAllData.csv

```
## # A tibble: 5 x 81
##   Date     BirdID GRP   Season AgeMeasD Sex   NestID BirdSeas GrpSizeAD X.Imm
##   <chr>    <chr>  <chr> <chr>     <int> <chr> <chr>  <chr>    <chr>     <chr>
## 1 3-Oct-05 MLYG   HAR   C            11 F     HARN1C MLYGC    6         0
## 2 3-Oct-05 PMPT   HAR   C            11 F     HARN1C PMPTC    6         0
## 3 3-Oct-05 TRMB   HAR   C            11 M     HARN1C TRMBC    6         0
## 4 4-Oct-05 WRMG   XHO   C            13 F     XHON1C WRMGC    6         2
## 5 4-Oct-05 RMPR   XHO   C            13 M     XHON1C RMPRC    6         2
## # ... with 71 more variables: GrpSizeTotal <chr>, TwoMonthsPrior <chr>,
## #   IncDate <chr>, HatchDate <chr>, FledgeDate <chr>, FailDate <chr>,
## #   IndepDate <chr>, X1YrDate <chr>, X2YrDate <chr>, BroodSize <int>,
## #   NoMales <dbl>, BroodSexRatio <dbl>, No.ChicksFledge <dbl>,
## #   MaleFledge <dbl>, NoSurv <dbl>, MaleSurv <dbl>, Mass <dbl>, LTars <dbl>,
## #   RTars <dbl>, RTarsMm <dbl>, DiffTars <dbl>, NestFledge <chr>,
## #   SurvFledge <int>, SurvInd <int>, SurvMassAtInd <dbl>, MassGainAtInd <dbl>,
## #   PropChangeMassAtInd <dbl>, Surv1Yr <int>, Surv1Mass <chr>,
## #   X1YrNatalGrp <int>, TMaxMeas <dbl>, Tmaxcat <chr>, MeanTmaxInc <dbl>,
## #   MeanTmaxBrood <dbl>, MeanTmaxNest <dbl>, MeanTmaxInd <dbl>, TMinMeas <dbl>,
## #   MeanTminInc <dbl>, MeanTminBrood <dbl>, MeanTminNest <dbl>,
## #   MeanTmin90 <dbl>, TvarMeas <dbl>, MeanTvarInc <dbl>, MeanTvarBrood <dbl>,
## #   MeanTvarNest <dbl>, MeanTvar90 <dbl>, HWaveInc <int>, NoHWInc1 <int>,
## #   NoHWInc2 <int>, NoHotDaysInc <int>, PropHotDaysInc <dbl>, HWaveBr <int>,
## #   NoHWBr1 <int>, NoHWBr2 <int>, NoHotDaysBr <int>, PropHotDaysBr <dbl>,
## #   HWaveNest <int>, NoHWNest1 <int>, NoHWNest2 <int>, NoHotDaysNest <int>,
## #   PropHotDaysNest <dbl>, NoHotDays90 <int>, PropHotDays90 <dbl>,
## #   RainTwoMonthsPrior <dbl>, Rain90 <dbl>, TotRain <dbl>, Drought <int>,
## #   X <lgl>, X.1 <int>, X.2 <lgl>, X.3 <lgl>
```

A preview of the dataset named Data.Bourne_FledglingSurvival.csv

```
## # A tibble: 5 x 39
##   Date      BirdID Group Season AgeMeasD Sex   NestID GrpSizeAD GrpSizeTotal
##   <chr>     <chr>  <chr> <chr>     <int> <chr> <chr>      <int>        <int>
## 1 03-Oct-05 MLYG   HAR   C            11 F     HARN1C         6            6
## 2 03-Oct-05 PMPT   HAR   C            11 F     HARN1C         6            6
## 3 04-Oct-05 WRMG   XHO   C            13 F     XHON1C         6            8
## 4 04-Oct-05 RMPR   XHO   C            13 M     XHON1C         6            8
## 5 12-Oct-05 PPRM   TPT   C            11 F     TPTN1C         7            7
## # ... with 30 more variables: PairTenure <int>, BroodSize <int>, NoMales <dbl>,
## #   BroodSexRatio <dbl>, Mass <dbl>, SurvInd <int>, SurvPeriod <int>,
## #   SurvMassAtInd <dbl>, Surv1Yr <int>, Surv1Mass <chr>, MeanTmaxInc <dbl>,
## #   MeanTmaxBrood <dbl>, MeanTmaxInd <dbl>, MeanTminInc <dbl>,
## #   MeanTminBrood <dbl>, MeanTmin90 <dbl>, MeanTvarInc <dbl>,
## #   MeanTvarBrood <dbl>, MeanTvar90 <dbl>, HWaveInc <int>, NoHotDaysInc <int>,
## #   PropHotDaysInc <dbl>, HWaveBr <int>, NoHotDaysBr <int>,
## #   PropHotDaysBr <dbl>, NoHotDays90 <int>, PropHotDays90 <dbl>,
## #   RainTwoMonthsPrior <dbl>, Rain90 <dbl>, Drought <int>
```

A preview of the dataset named Data.Bourne_NestSuccess.csv:

```
## # A tibble: 5 x 38
##   Group Season NestCode IncPeriod IncPeriodHatched~ BroodPeriod BroodPeriodFled~
##   <chr> <chr>  <chr>        <int>             <int>       <int>            <int>
## 1 RNB   P      RNBN1P          16                16          14               14
## 2 RNB   P      RNBN2P          14                14          15               15
## 3 INF   Q      INFN2Q          13                13           4               NA
## 4 INF   Q      INFN3Q           4                NA          NA               NA
## 5 BAL   C      BALN3C          14                14          16               16
## # ... with 31 more variables: Period90 <int>, TotPerCutOff <int>, Event <int>,
## #   HatchCatNum <int>, EventInc <int>, FledgeCatNum <int>, EventBr <int>,
## #   FledgeExclNotHatched <int>, NoChicksHatched <dbl>, NoChicksFledged <dbl>,
## #   MeanTmaxInc <dbl>, MeanTmaxBrood <dbl>, GrpSizeAD <int>,
## #   RainfallTwoMonthsPrior <dbl>, PropHotDaysInc <dbl>, PropHotDaysBr <dbl>,
## #   MeanTmax90 <dbl>, PropHotDays90 <dbl>, Rain90 <dbl>, PairTenure <int>,
## #   LongestDom <int>, AveTmax <dbl>, AvePropHot <dbl>, TotRain <dbl>,
## #   SurvInd <int>, MeanTminInc <dbl>, MeanTminBr <dbl>, MeanTmvarInc <dbl>,
## #   MeanTvarBr <dbl>, MeanTmin90 <dbl>, MeanTvar90 <dbl>
```

Before merging the Nestling and Fledgling dataframes, several actions need to be taken to cleanse the data and make it ready. They are detailed below:

Nestling data

- Convert all date related fields from character to date format
- Remove any potential whitespace from character fields
- Convert Surv1Mass,GrpSizeAD,GrpSizeTotal and X.Imm to numeric type from character.
- Renaming GRP to Group

Fledgling data

- Convert date related field to date format
- Convert Surv1Mass to numeric
- Remove whitespaces from character fields

It does appear from NAs were introduced during the data conversion of characater to numeric, which are for fields that did not have any data to start with.

The dimensions of the merged data set are 596 rows and 83 columns. The nestling data was used as the base as this is the earlier developmental stage before they become fledglings.Some birds do not make it to fledgling thus the nestling data has more rows/birds. Therefore, the fledgling data is left-joined to nestlings to get the most complete picture possible.

Part 2

2.1)The Total Sample Size Per Year:

Table 1: The Total Sample Size Per Year

year	Sample Size
2005	39
2006	60
2007	28
2008	58
2009	33
2010	42
2011	56
2012	45
2013	46
2014	74
2015	30
2016	25
2017	42
2018	18

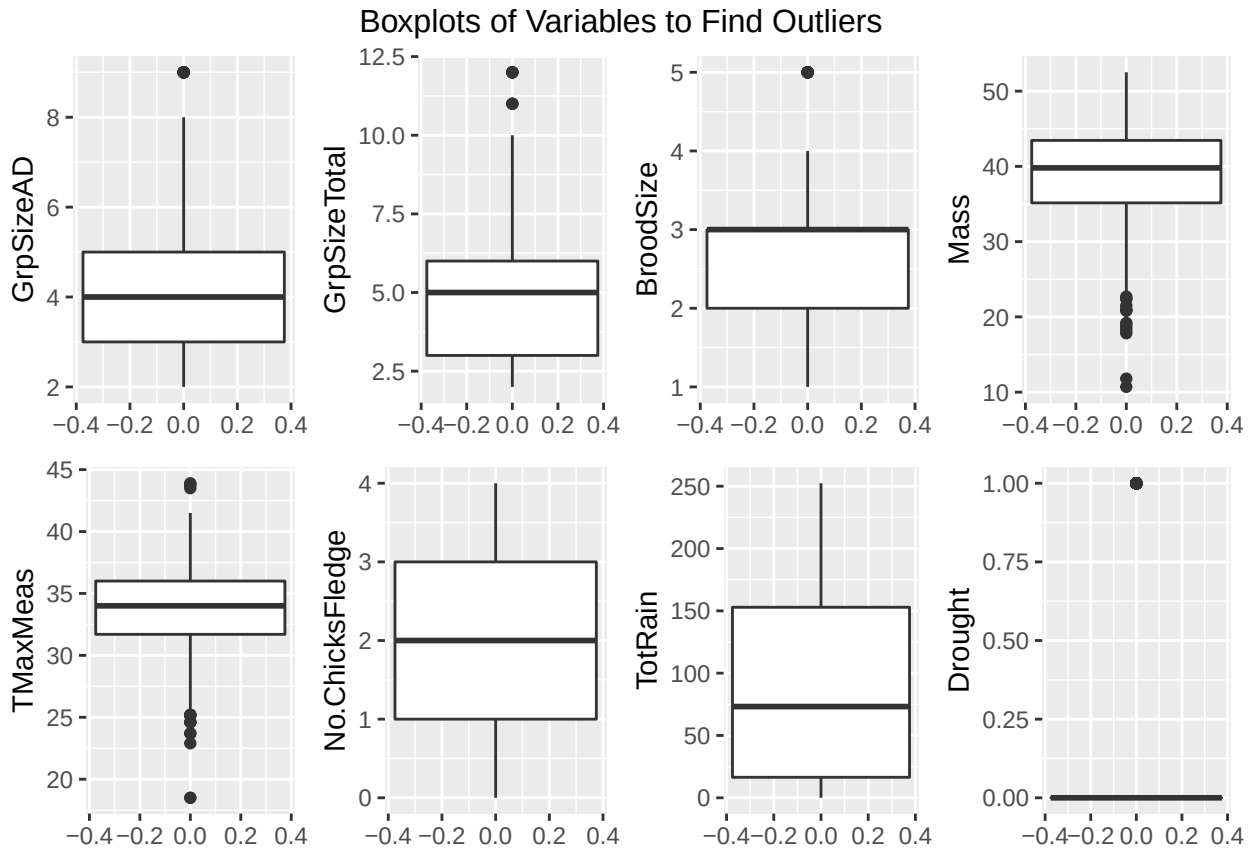
2.2.1) Checking for Nulls

Table 2: Number of Missing Values in Selected Fields

	Count
GrpSizeAD	1
GrpSizeTotal	1
BroodSize	0
Mass	5
TMaxMeas	37
No.ChicksFledge	1
TotRain	26
Drought	26

The occurance of nulls indicates we need to be cater for them when plotting or modeling with these fields.

2.2.2)Boxplots to allow us to look for outliers:



An outlier is an observation that lies at or further away than 1.5 times the interquartile range, either below the lower quartile or above the upper quartile i.e. any observations further out than the "whiskers' on a box plot.

The following variables appear to have 1 or more outliers:

- GrpSizeAD
- GrpSizeTotal
- BroodSize
- Mass
- TMaxMeas

The following variable appear to not have outliers:

- No.ChicksFledge
- TotRain

The variable Drought appears to be a binary indicator (1/0) as to whether a drought occurred, and thus would not have any outliers; despite what the boxplot appears to indicate "1" is a valid observation and not an outlier

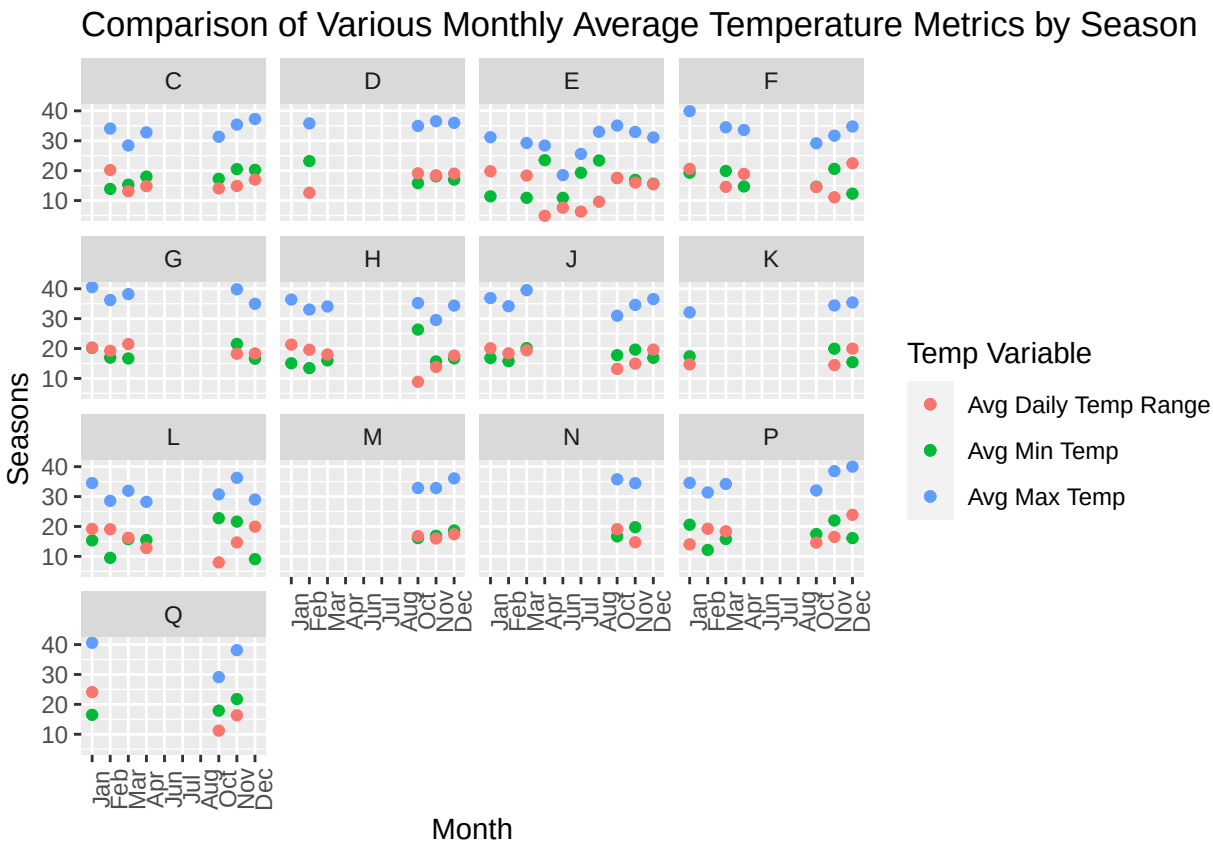
2.3)

Table 3: Preview of Mass and Date per Nest

NestID	Date	Total Mass
AQAN1C	2005-10-22	77.1
AQAN1D	2006-10-14	74.7
AQAN2C	2005-12-19	23.0
BABN1D	2006-11-08	65.2
BABN3C	2005-11-28	63.7

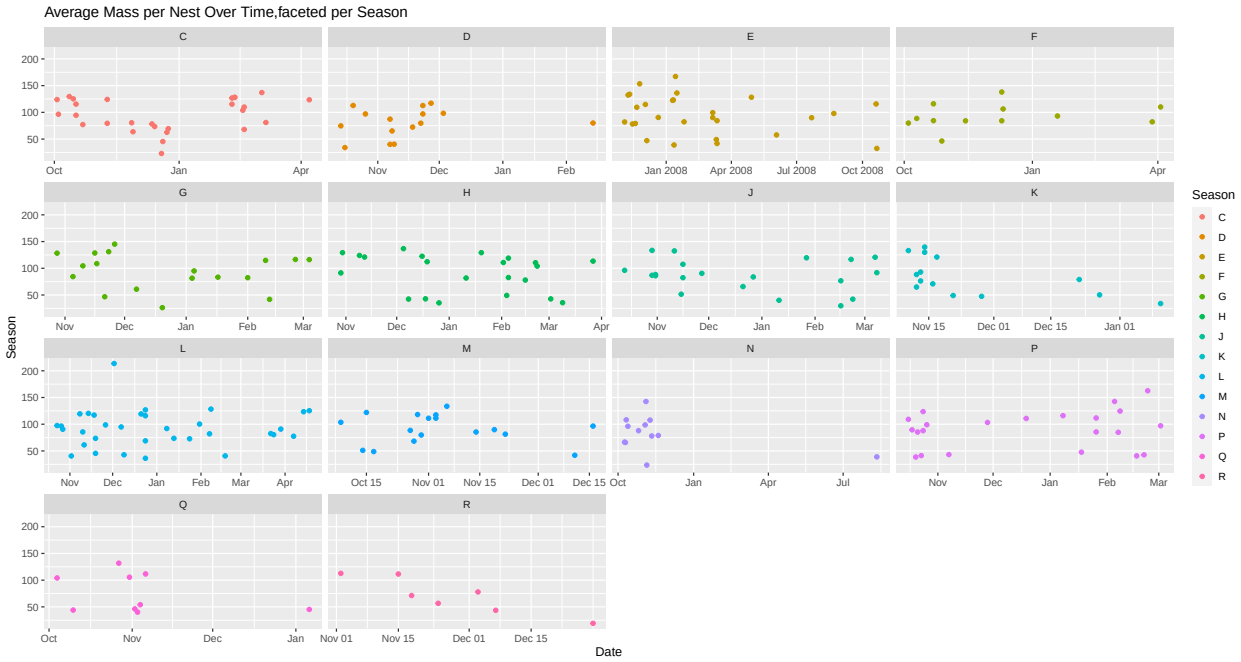
The dimensions of the table are 257 rows and 3 columns.

2.4)



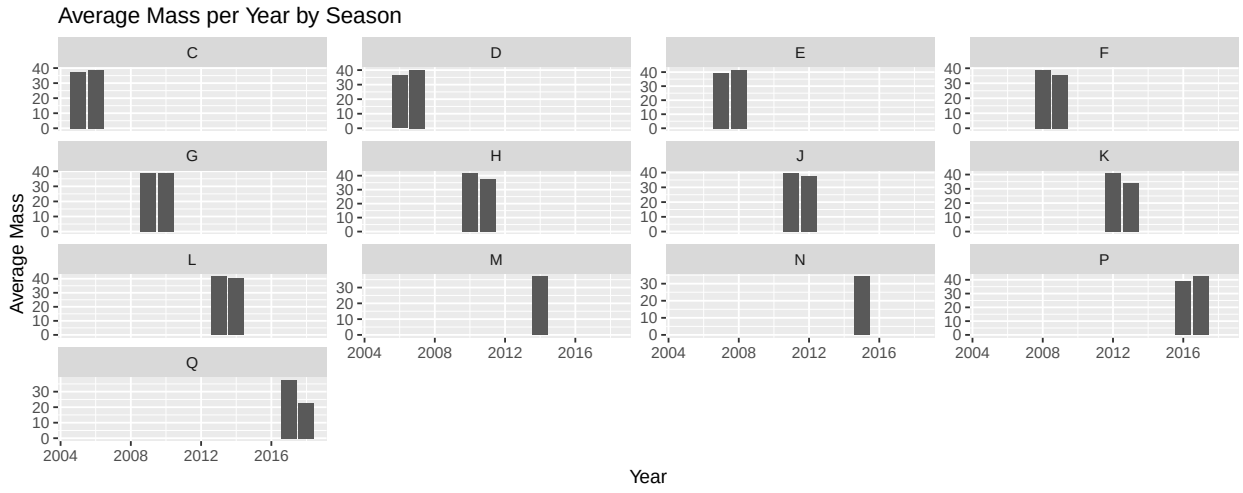
The figure above shows us how the average temperature range, the average maximum temperature and the average minimum temperature changed on a monthly basis for different seasons. This gives us an of whether one season had more extreme temperatures and temperature variance when compared to another.

2.5)

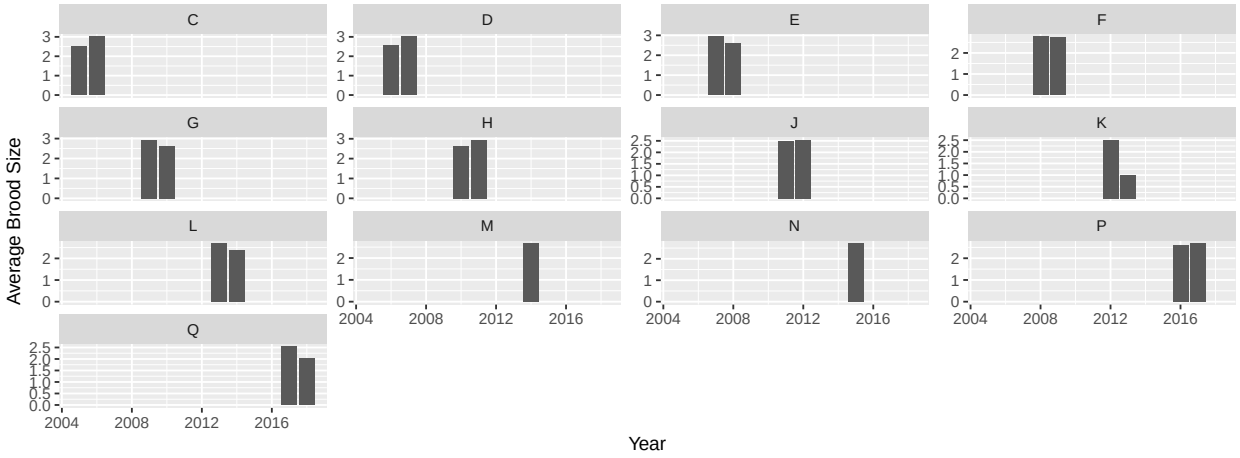


A quick overview does not reveal any obvious trends of how the total mass of nests may have changed over time.

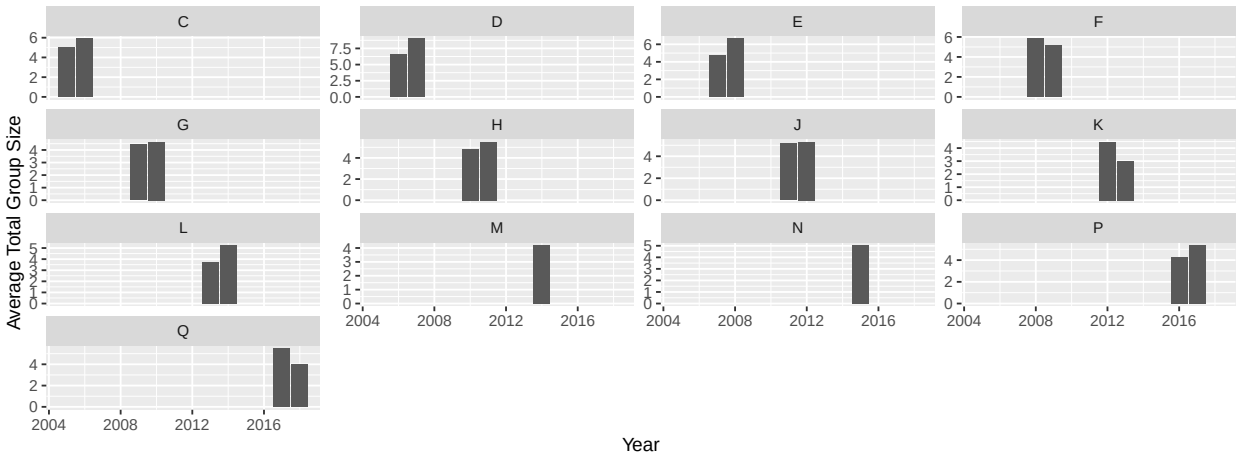
2.6)



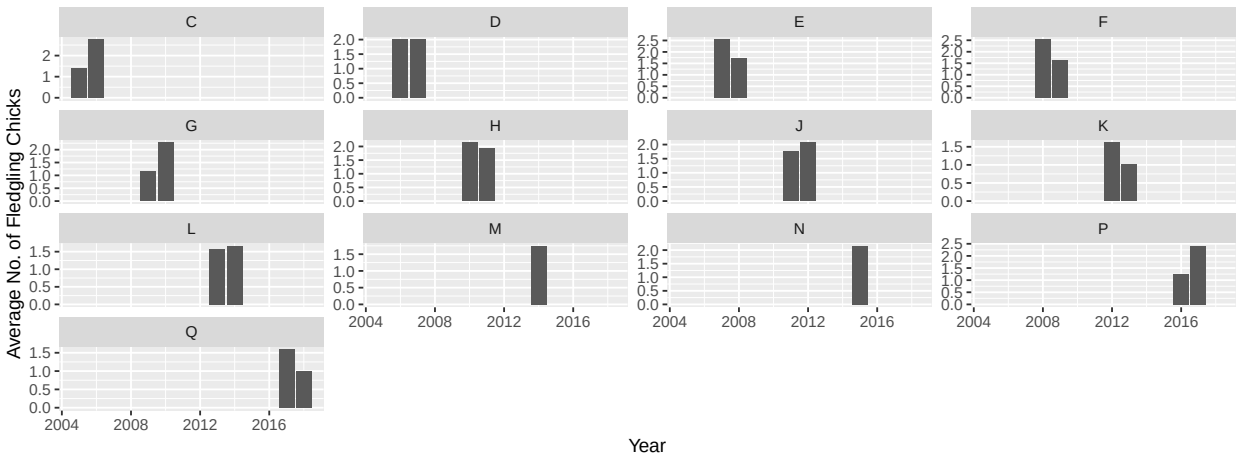
Average Brood Size per Year by Season

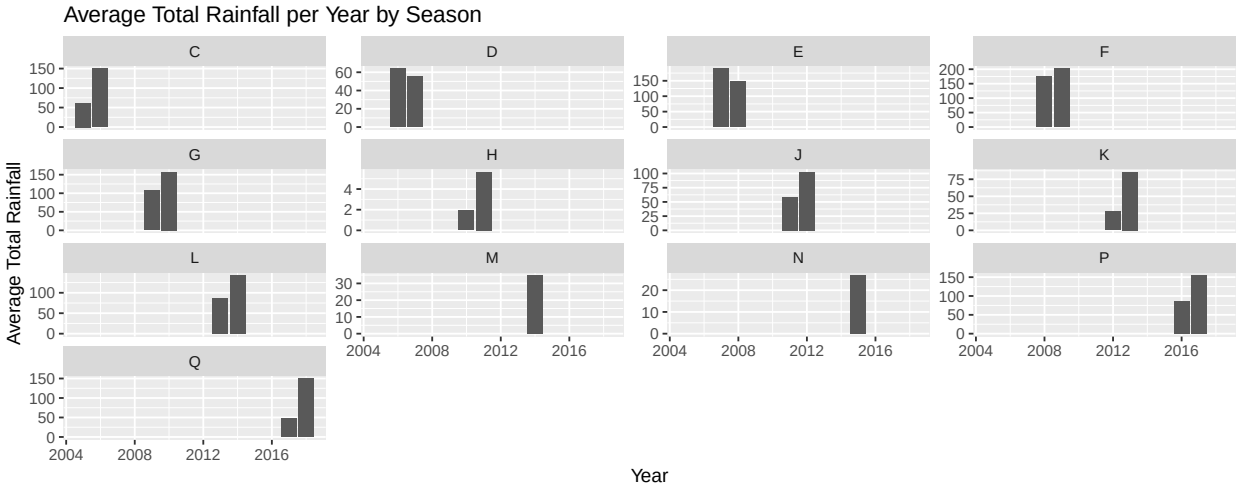


Average Total Group Size per Year by Season

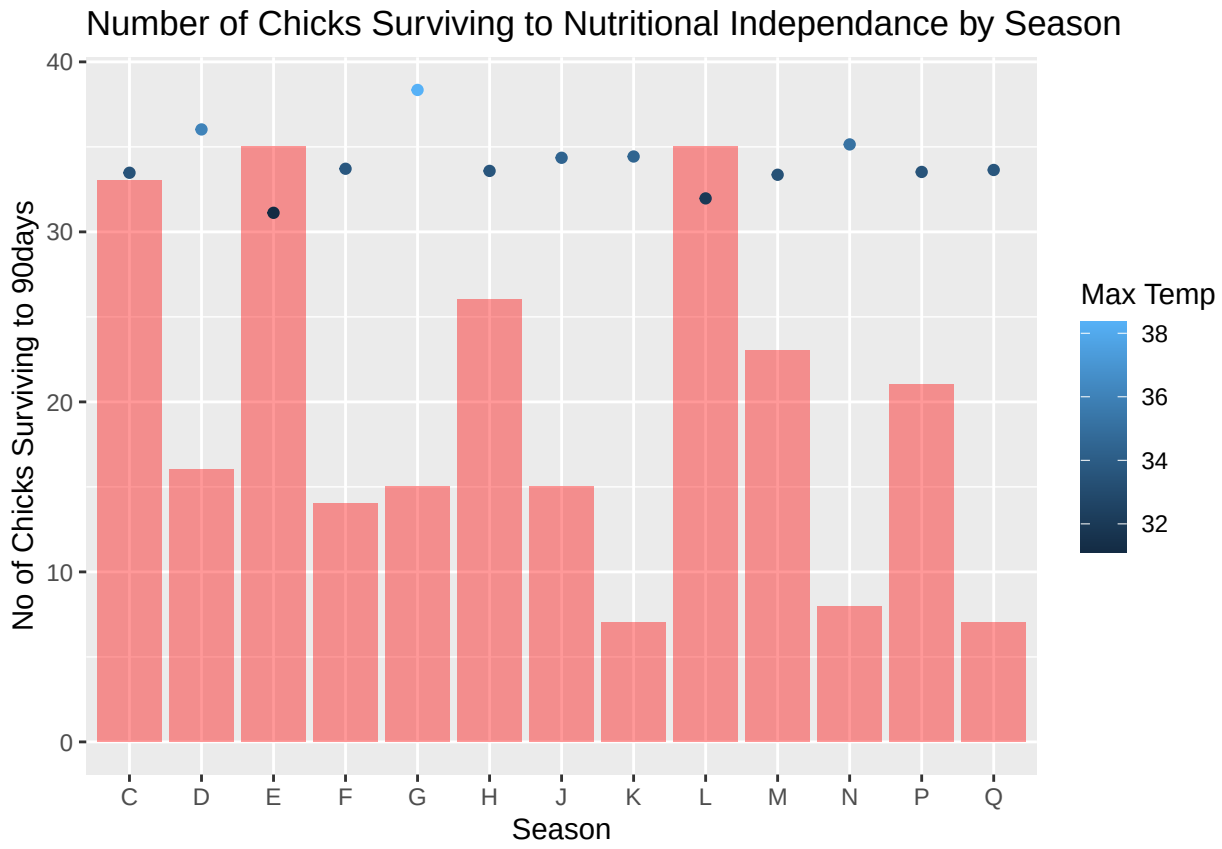


Average No. of Fledgling Chicks per Year by Season





2.7) The study this data was collected for was investigating if the survivability of the young birds at different developmental stages is affected by environmental conditions and/or group size. Therefore it makes sense to plot a graphic with an environmental factor such as temperature against the number of young that survive.



From the figure above, it does appear that higher maximum temperatures could affect the survivability of young birds. The two seasons with the highest survivability numbers are those with the lowest maximum temperature. However, this is not concrete enough evidence and warrants further investigation and testing.