# ST117 Individual DRAFT Written Report - Part Birds

My WARWICK ID 5538165 (Report Pod 017)

Today's date in the format 2024-04-17

- Please read the document "WR Assessment, Style, and Submission Guidance".
- The upper limits given below are not relevant for your individual DRAFT submission. We only mention them here in view of your work on the final report. Your individual DRAFT report may be much shorter. (See guidance document regarding the rationale for the individual DRAFT report.)
- Please replace 1234567, 099, and 1999-12-31 in the header by the appropriate numbers.

## Question 1 (Polynomial regression)

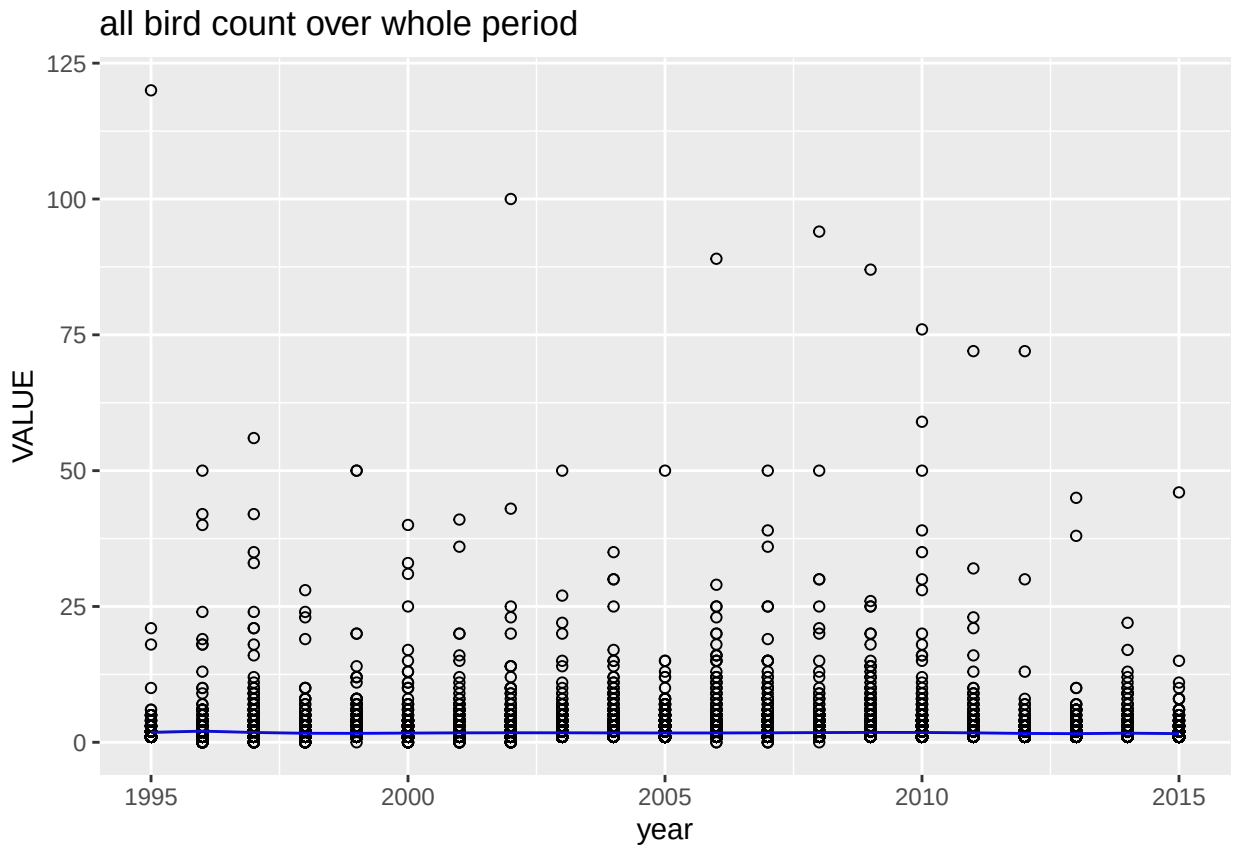
### a) for all species in all locations

***Upper limit: 2 figures, 1 table, 50 words***

```
# import bird data and add a new date and year column
Raw_bird_data <- read.csv("../../ST117 Project/00_raw_data/Bird Raw Data.csv")
Raw_bird_data$date <- dmy(Raw_bird_data$SDATE)
Raw_bird_data$year <- year(Raw_bird_data$date)

# we then use the same code/function we used for the moth questions
fit = lm(VALUE ~ poly(year, 10), data = Raw_bird_data)
yearlims = range(Raw_bird_data$year)
year_grid = seq(from = min(yearlims), to = max(yearlims))
cat("year: ", year_grid)
```

```
## year:  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
```

```
preds = predict(fit, newdata = list(year = year_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = Raw_bird_data, pch=1, aes(x = year, y = VALUE)) +
  geom_line(aes(x = year_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = year_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(yearlims) +
  labs(title = paste("all bird count over whole period "))
```




b) for your report pod's species in all locations

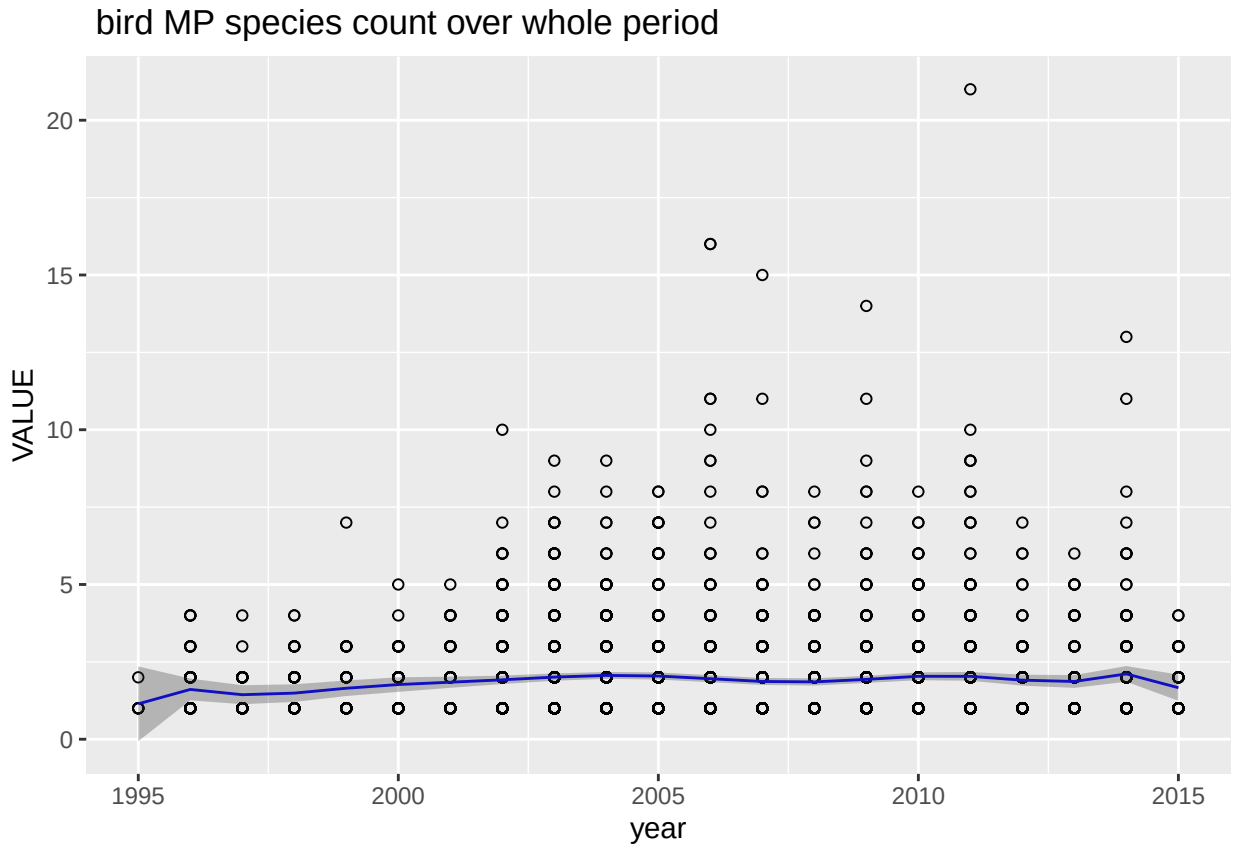
***Upper limit: 8 figures (e.g., for each of the 4 species: 2 figures), 1 table, 100 words***

```
# we now filter the raw data into 4 separate data frames for each species
bird_MP <- filter(Raw_bird_data,FIELDNAME == "MP")
bird_WR <- filter(Raw_bird_data,FIELDNAME == "WR")
bird_S <- filter(Raw_bird_data,FIELDNAME == "S")
bird_C <- filter(Raw_bird_data,FIELDNAME == "C")
# same code again for polynomial regression models for each species

fit = lm(VALUE ~ poly(year, 10), data = bird_MP)
yearlims = range(bird_MP$year)
year_grid = seq(from = min(yearlims), to = max(yearlims))
cat("year: ", year_grid)
```

```
## year:  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
```

```
preds = predict(fit, newdata = list(year = year_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
```

```
ggplot() +
  geom_point(data = bird_MP, pch=1, aes(x = year, y = VALUE)) +
  geom_line(aes(x = year_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = year_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(yearlims) +
  labs(title = paste(" bird MP species count over whole period "))
```

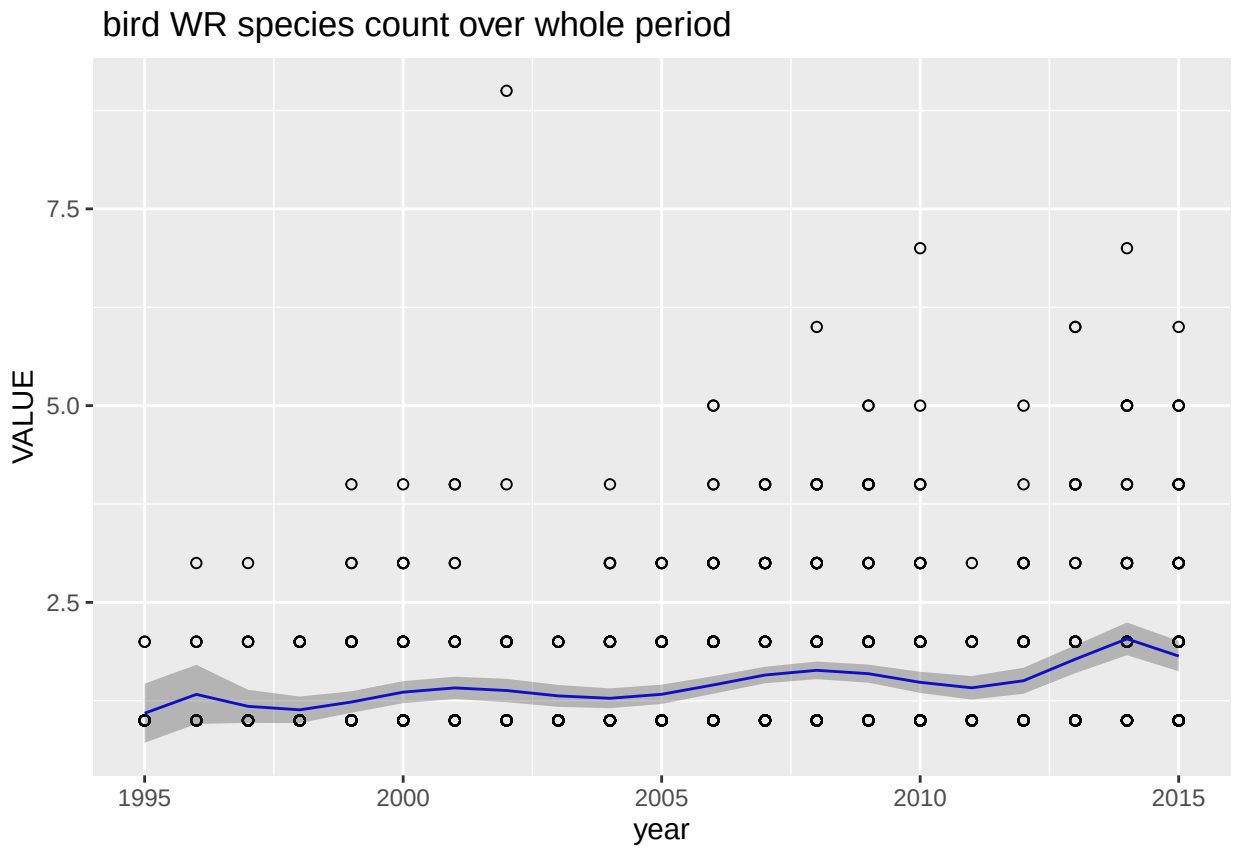



```
fit = lm(VALUE ~ poly(year, 10), data = bird_WR)
yearlims = range(bird_WR$year)
year_grid = seq(from = min(yearlims), to = max(yearlims))
cat("year: ", year_grid)
```

```
## year:  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
```

```
preds = predict(fit, newdata = list(year = year_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = bird_WR, pch=1, aes(x = year, y = VALUE)) +
  geom_line(aes(x = year_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = year_grid,
```

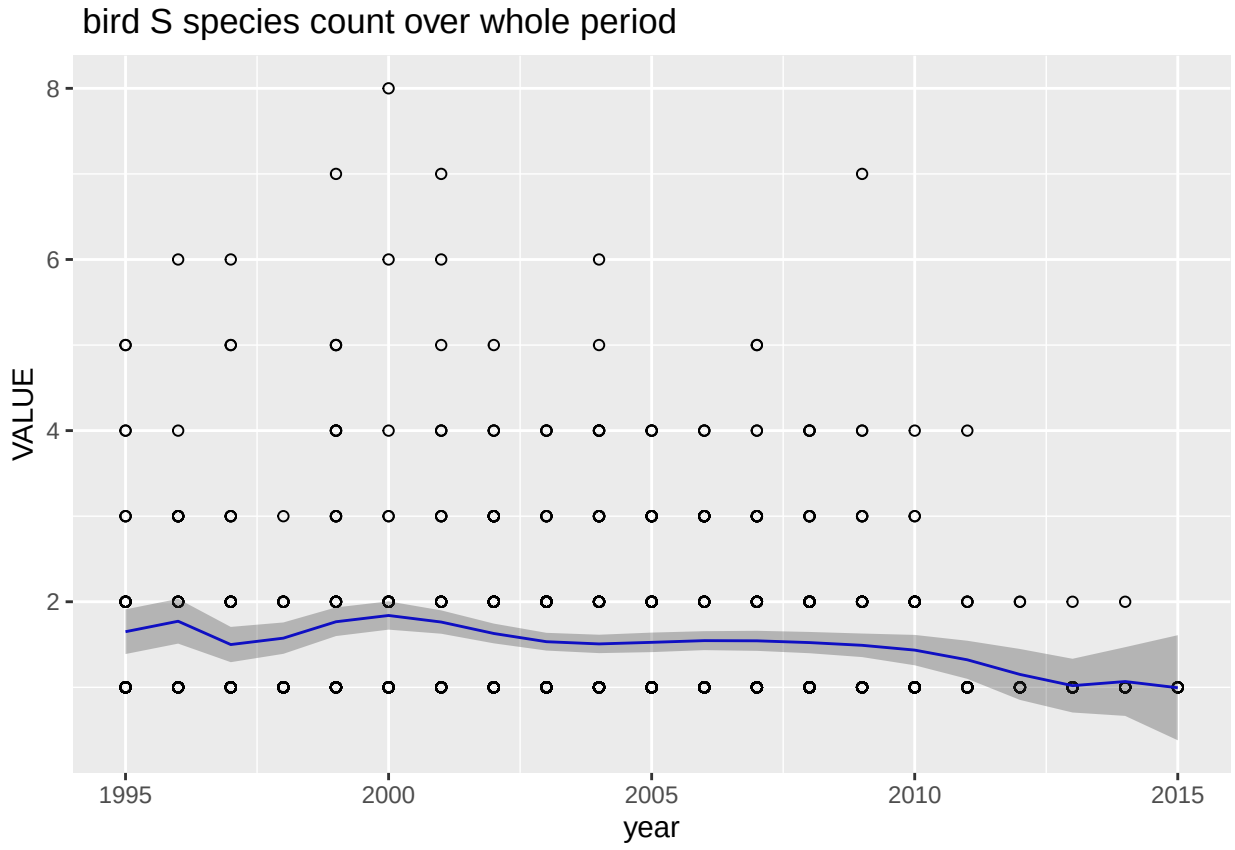
```
                ymin = se_bands[,"lower"],
                ymax = se_bands[,"upper"]),
            alpha = 0.3) +
  xlim(yearlims) +
  labs(title = paste(" bird WR species count over whole period "))
```




```
fit = lm(VALUE ~ poly(year, 10), data = bird_S)
yearlims = range(bird_S$year)
year_grid = seq(from = min(yearlims), to = max(yearlims))
cat("year: ", year_grid)
```

```
## year:  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
```

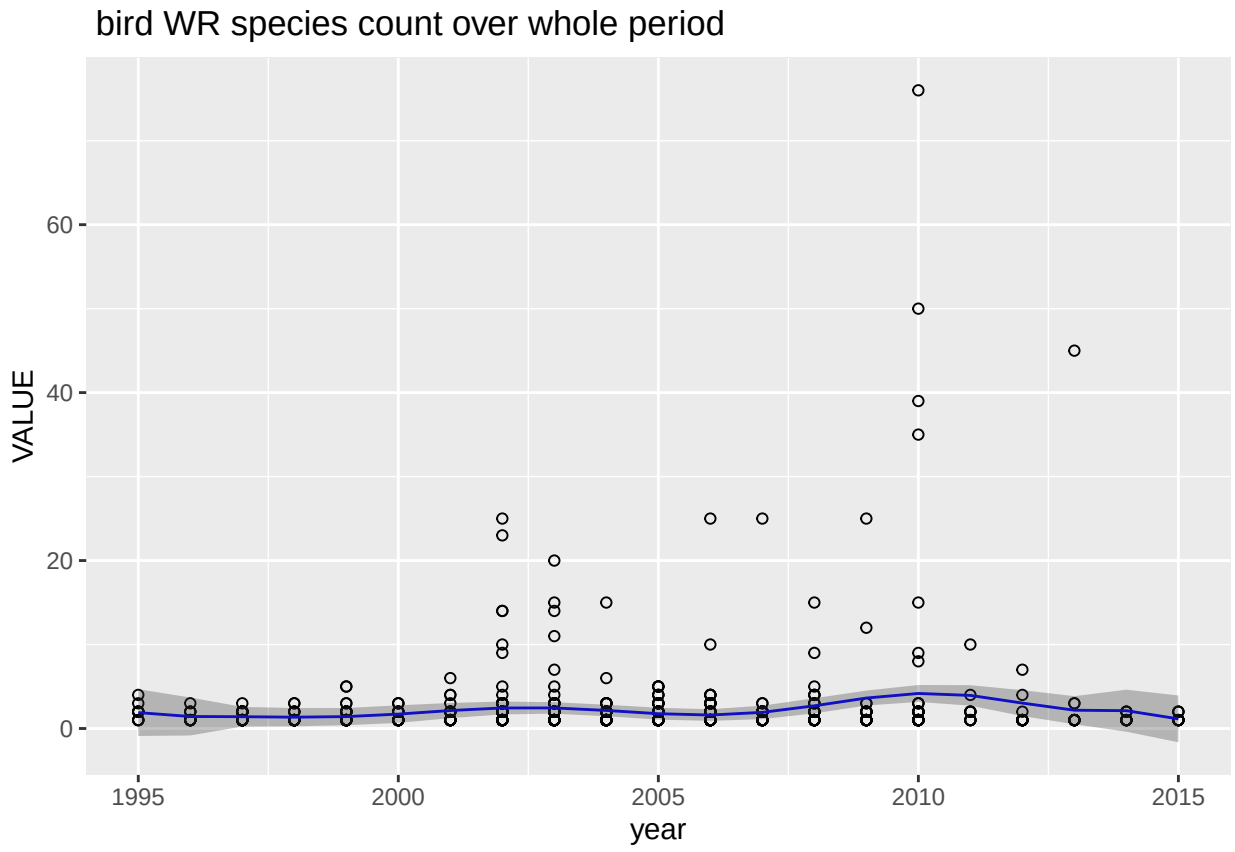
```
preds = predict(fit, newdata = list(year = year_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = bird_S, pch=1, aes(x = year, y = VALUE)) +
  geom_line(aes(x = year_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = year_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(yearlims) +
  labs(title = paste(" bird S species count over whole period "))
```




```
fit = lm(VALUE ~ poly(year, 10), data = bird_C)
yearlims = range(bird_C$year)
year_grid = seq(from = min(yearlims), to = max(yearlims))
cat("year: ", year_grid)
```

```
## year:  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
```

```
preds = predict(fit, newdata = list(year = year_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = bird_C, pch=1, aes(x = year, y = VALUE)) +
  geom_line(aes(x = year_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = year_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(yearlims) +
  labs(title = paste(" bird WR species count over whole period "))
```




### c) for all species in the locations with the highest overall bird count

***Upper limit: 2 figures, 1 table, 50 words***

Start by stating what is the location in question.

```
# start by making a sum function for each location to find which location has the most birds

sumofbird_bysitecode <- Raw_bird_data %>%
  group_by(SITECODE) %>%
  summarise(sumbirds = sum(VALUE))

print(sumofbird_bysitecode)
```

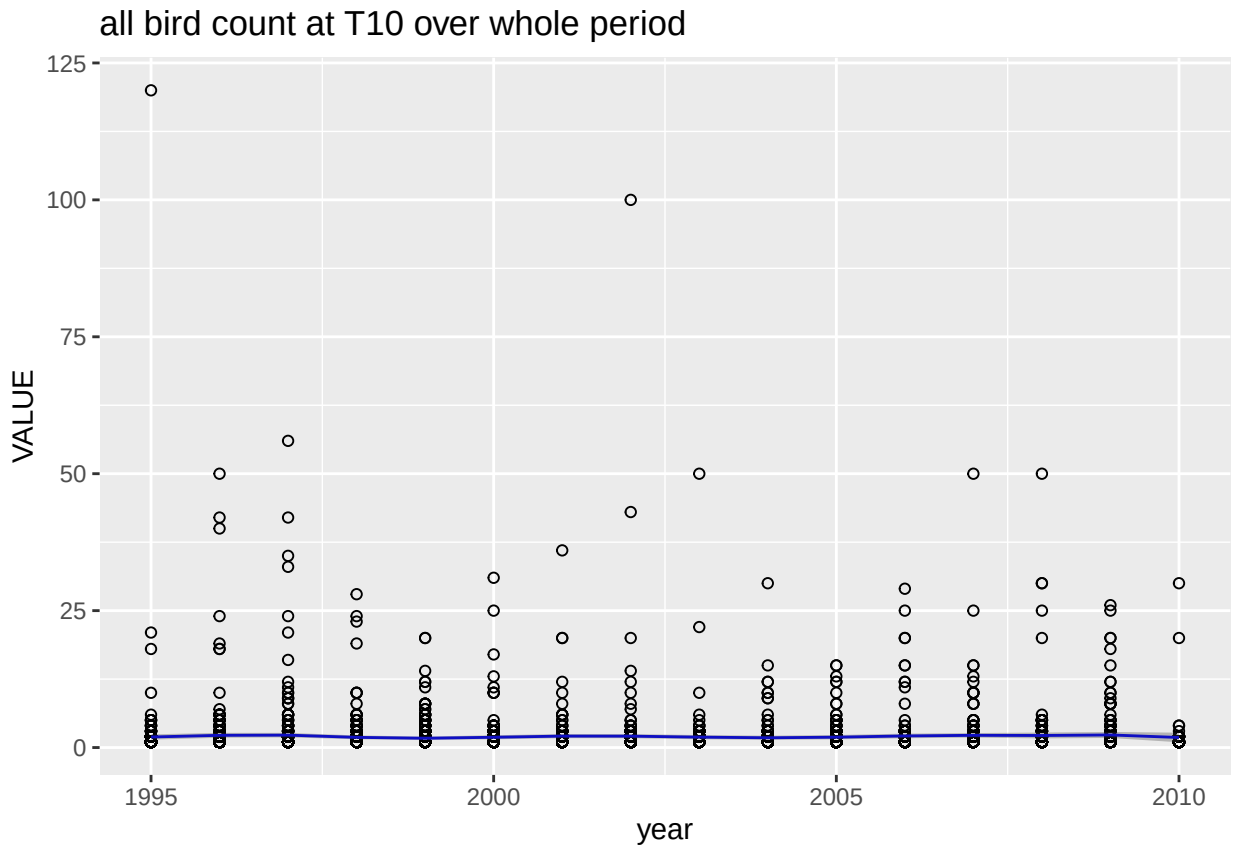
```
## # A tibble: 11 x 2
##    SITECODE sumbirds
##    <chr>       <int>
##  1 T01          2974
##  2 T02          3737
##  3 T03          4614
##  4 T04          2214
##  5 T05          4144
##  6 T06          3391
##  7 T07          4097
```

```
##  8 T09          2473
##  9 T10          8419
## 10 T11          1472
## 11 T12          1596
```

```r
# now filter the data with T10 sitecode

bird_T10_data <- filter(Raw_bird_data, SITECODE =="T10")

# plot the model for this data

fit = lm(VALUE ~ poly(year, 10), data = bird_T10_data)
yearlims = range(bird_T10_data$year)
year_grid = seq(from = min(yearlims), to = max(yearlims))
cat("year: ", year_grid)
```

```
## year:  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
```

```r
preds = predict(fit, newdata = list(year = year_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = bird_T10_data, pch=1, aes(x = year, y = VALUE)) +
  geom_line(aes(x = year_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = year_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(yearlims) +
  labs(title = paste("all bird count at T10 over whole period "))
```




T10 is the most frequent location as we can see from the table.

### Conclusion

Summarise what the models suggest about development of ECN bird counts over the study period ## *Upper limit: 100 words*

## Question 2

### Figure A

#### Explain plot layout

***Upper limit: 80 words***

As we can see this plot has many box plots that are parallel to each other. The coloured circles match each sitecode (location) and the x axis seems to match each year of data collection. The y axis is the bird count per year.

### Describe what the plot shows about the ECN bird data

***Upper limit: 80 words***

The plot shows that roughly over the whole time period the median of the bird count per year for every sitecode is around the same however in the first year of data collection it was the greatest of all years. This could be due to the fact that it was the first year of data collection hence there could be errors. Up to 2009, T10 is always near the greatest bird count per year out of all the sitecodes however after 2010 T10 decreases and T03 begins to increase as T10 decreases. This could indicate a change in the environment since the start of data collection as T10's environment may have been affected so that birds die or move away.

### Missing annotation

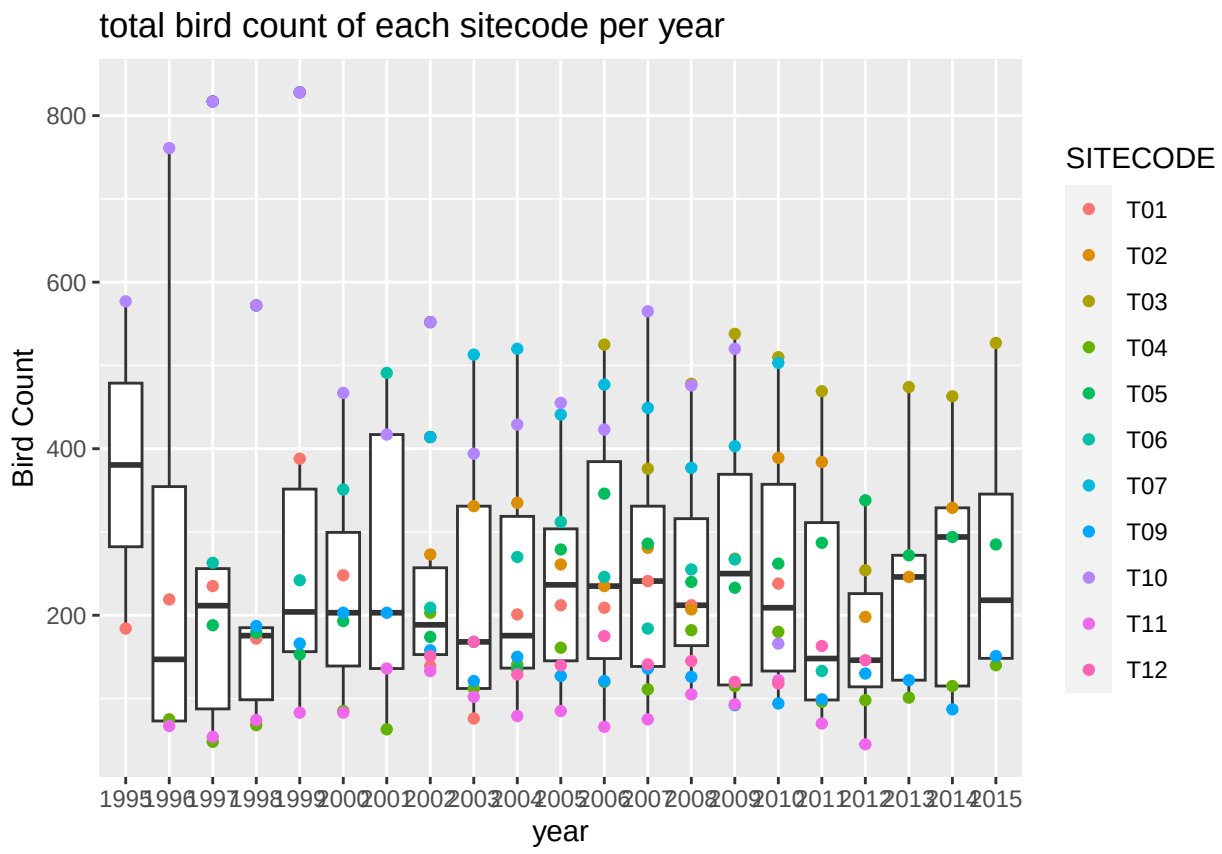
- axes labels:
- tick marks:
- title:
- legend labels: These are shown on the plot.

### Reengineer R code and output plot

```
# we need to find the total bird count for each year since data collection
# but we also need to get this value for each location
year_sitecode_bird <- Raw_bird_data%>%
  group_by(year,SITECODE)%>%
  summarise(total_bird_count=sum(VALUE))
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
# now we can make parallel box plots using colour=SITECODE to ensure the sitecodes are distinguishable,
ggplot(data = year_sitecode_bird,aes(x=factor(year),y=total_bird_count)) +
  geom_boxplot()+
  geom_point(aes(colour=SITECODE))+
  labs(title = "total bird count of each sitecode per year", x = "year", y = "Bird Count")
```




## Figure B

### Explain plot layout

***Upper limit: 80 words***

Similar to the plot above this is of same format however the box plots are for each sitecode rather than for each year. The y axis is made up of the column from the year_sitecode_bird data frame and this column represents the sum of the bird count for each year and sitecode.

### Describe what the plot shows about the ECN bird data

***Upper limit: 80 words***

The plot shows us that certain sitecodes such as T11 and T12 have had similar bird counts per year over the whole period however locations such as T03 and T10 have years that look to be anomalous. As we can see for T10 in the early years (around 1995) we have a very high bird count ~800 but then as the years increase we see it drop to ~150 around 2015. This plot enables us to see how much certain locations have been affected by climate change from 1995-2015.

### Missing annotation

- axes labels:

- tick marks:
- title:
- legend labels: These are on the plot ### Reengineer R code and output plot

```
# we now use the same data frame we made in part a but we swap the SITECODE with the year column.
ggplot(data = year_sitecode_bird,aes(x=factor(SITECODE),y=total_bird_count)) +
  geom_boxplot()+
  geom_point(aes(colour=year))+
  labs(title = "total bird count of each year per sitecode", x = "SITECODE", y = "Bird Count")
```