ST117 Individual DRAFT Written Report - Part Moths

My WARWICK ID 5538165 (Report Pod 017)

Today's date in the format 2024-04-16

- To make your report easier to read you may group figures corresponding to a unit of analysis in a grid. Example code chunks for creating such layouts are given below.

- 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

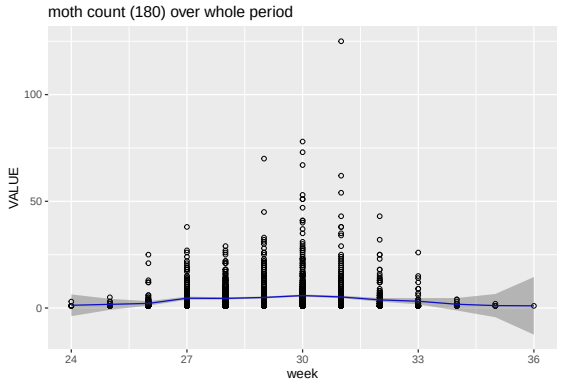
a)

Upper limit: 10 figures (e.g., for each of the 5 species: 2 figures), 1 table, 150 words

```
# download data then set start date and end date to then give the data frame a new
# column with week number the date is in
raw_moth_data <- read.csv("../../ST117 Project/00_raw_data/Moth Raw Data.csv")
# give data a date column like in the EDA
raw_moth_data$date <- dmy(raw_moth_data$SDATE)
start_date <- min(raw_moth_data$date)
end_date <- max(raw_moth_data$date)
raw_moth_data$week <- as.numeric(format(raw_moth_data$date, "%V"))
raw_moth_data$year <- year(raw_moth_data$date)

# filter data in the same way for each species of moth
filtered_moth_data302 <- filter(raw_moth_data,FIELDNAME == 302 )
filtered_moth_data180 <- filter(raw_moth_data, FIELDNAME == 180 )
filtered_moth_data382 <- filter(raw_moth_data, FIELDNAME == 382 )
filtered_moth_data648 <- filter(raw_moth_data, FIELDNAME == 648 )
filtered_moth_data313 <- filter(raw_moth_data,FIELDNAME == 313)

# from lectures use examples to give values to parameters to put in the
# polynomial regression plot code
fit = lm(VALUE ~ poly(week, 10), data = filtered_moth_data180)
weeklims = range(filtered_moth_data180$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  24 25 26 27 28 29 30 31 32 33 34 35 36
```

```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
#use calculated values to plot a polynomial regression

ggplot() +
  geom_point(data = filtered_moth_data180, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (180) over whole period "))
```



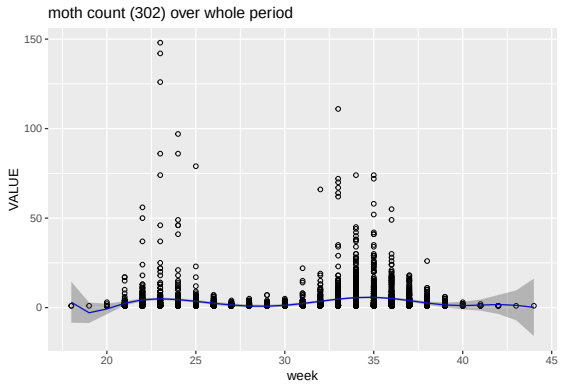
We then do the exact same with each of the 5 moth species hence I will not comment on this part of the code.

```
fit = lm(VALUE ~ poly(week, 10), data = filtered_moth_data302)
weeklims = range(filtered_moth_data302$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
```

```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = filtered_moth_data302, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
```

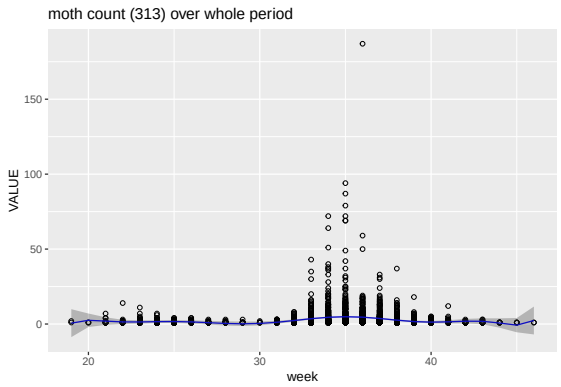
```
  xlim(weeklims) +
  labs(title = paste("moth count (302) over whole period "))
```




```
fit = lm(VALUE ~ poly(week, 10), data = filtered_moth_data313)
weeklims = range(filtered_moth_data313$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
```

```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = filtered_moth_data313, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (313) over whole period "))
```

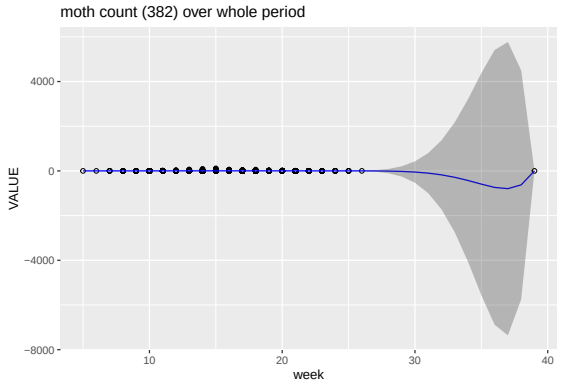



```
fit = lm(VALUE ~ poly(week, 10), data = filtered_moth_data382)
weeklims = range(filtered_moth_data382$week)
```

```
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
```

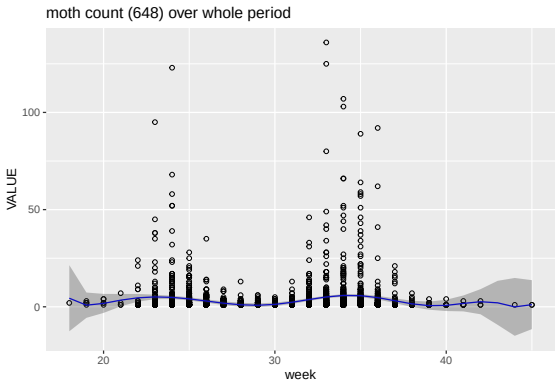
```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = filtered_moth_data382, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (382) over whole period "))
```




```
fit = lm(VALUE ~ poly(week, 10), data = filtered_moth_data648)
weeklims = range(filtered_moth_data648$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
```

```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = filtered_moth_data648, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (648) over whole period "))
```



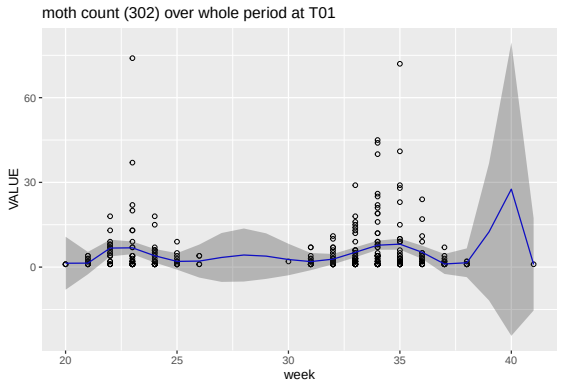
b)

Upper limit: 16 figures (e.g., for each of the 4 locations and 2 species: 2 figures), 1 table, 150 words

```
# filter the raw data down to the one species and for each of the 4 locations
moth_302_data <- filter(raw_moth_data, FIELDNAME == 302)
moth_302_T01 <- filter(moth_302_data,SITECODE =="T01")
moth_302_T02 <- filter(moth_302_data,SITECODE == "T02")
moth_302_T03 <- filter(moth_302_data,SITECODE =="T03")
moth_302_T04 <- filter(moth_302_data,SITECODE=="T04")

# We then use the same code to fit polynomial regression models as in part a
# Hence I will not comment on this code

fit = lm(VALUE ~ poly(week, 10), data = moth_302_T01)
weeklims = range(moth_302_T01$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
```

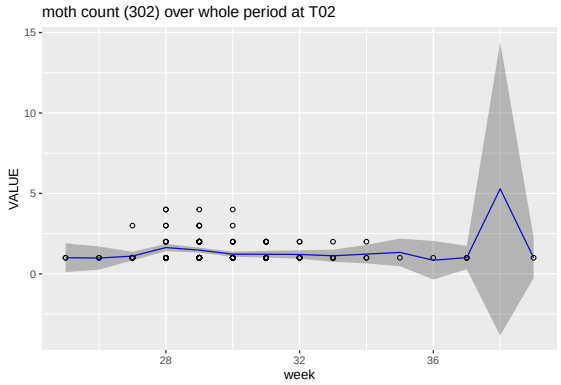
```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = moth_302_T01, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (302) over whole period at T01 "))
```




```
fit = lm(VALUE ~ poly(week, 10), data = moth_302_T02)
weeklims = range(moth_302_T02$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
```

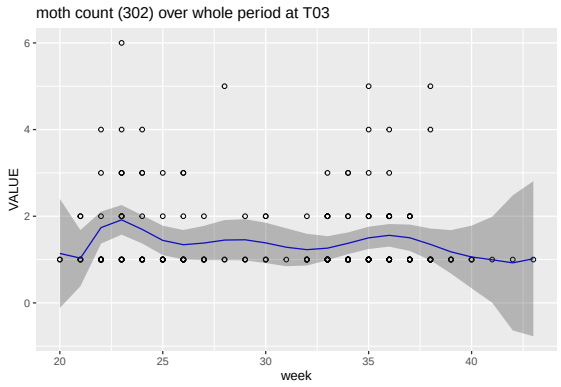
```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = moth_302_T02, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (302) over whole period at T02 "))
```




```
fit = lm(VALUE ~ poly(week, 10), data = moth_302_T03)
weeklims = range(moth_302_T03$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
```

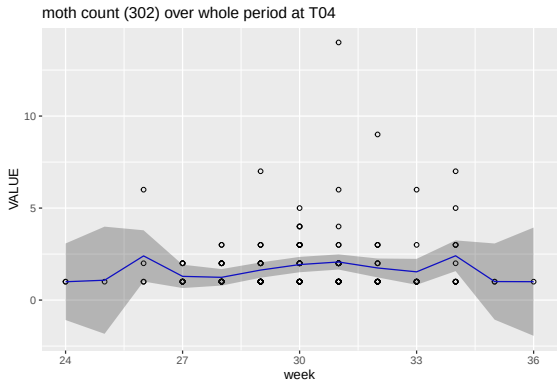
```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = moth_302_T03, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (302) over whole period at T03"))
```




```
fit = lm(VALUE ~ poly(week, 10), data = moth_302_T04)
weeklims = range(moth_302_T04$week)
week_grid = seq(from = min(weeklims), to = max(weeklims))
cat("week: ", week_grid)
```

```
## week:  24 25 26 27 28 29 30 31 32 33 34 35 36
```

```
preds = predict(fit, newdata = list(week = week_grid), se = TRUE)
se_bands = cbind("upper" = preds$fit+2*preds$se.fit,
                 "lower" = preds$fit-2*preds$se.fit)
ggplot() +
  geom_point(data = moth_302_T04, pch=1, aes(x = week, y = VALUE)) +
  geom_line(aes(x = week_grid, y = preds$fit), color = "#0000FF") +
  geom_ribbon(aes(x = week_grid,
                  ymin = se_bands[,"lower"],
                  ymax = se_bands[,"upper"]),
              alpha = 0.3) +
  xlim(weeklims) +
  labs(title = paste("moth count (302) over whole period at T04 "))
```



Conclusion

Summarise what the models suggest about the seasonality of the moths counts ## *Upper limit: 100 words*

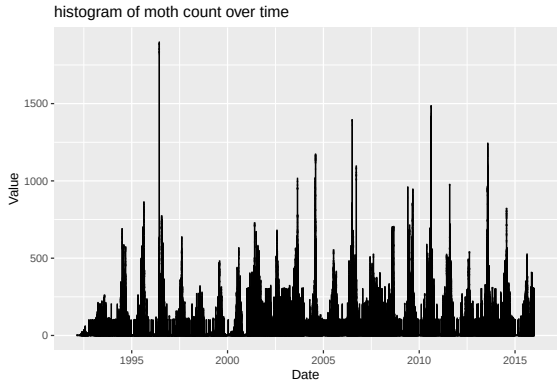
Question 2

a)

Upper limit: 2 figures, 100 words

```
# from the raw data we want moth count against time
ggplot(raw_moth_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moth count over time", x = "Date", y = "Value")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘
```

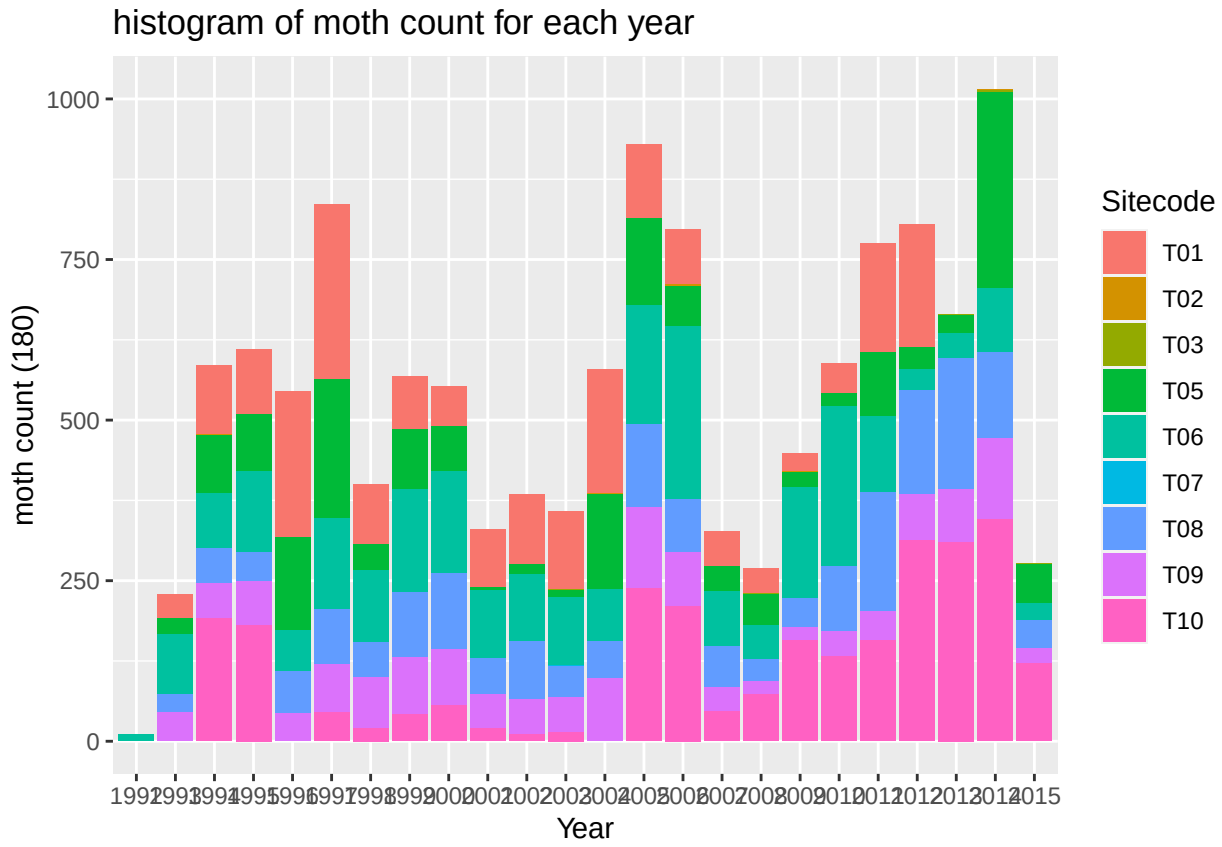


b)

Upper limit: 30 figures (e.g., for each of the 5 species: 6 figures), 1 table, 200 words

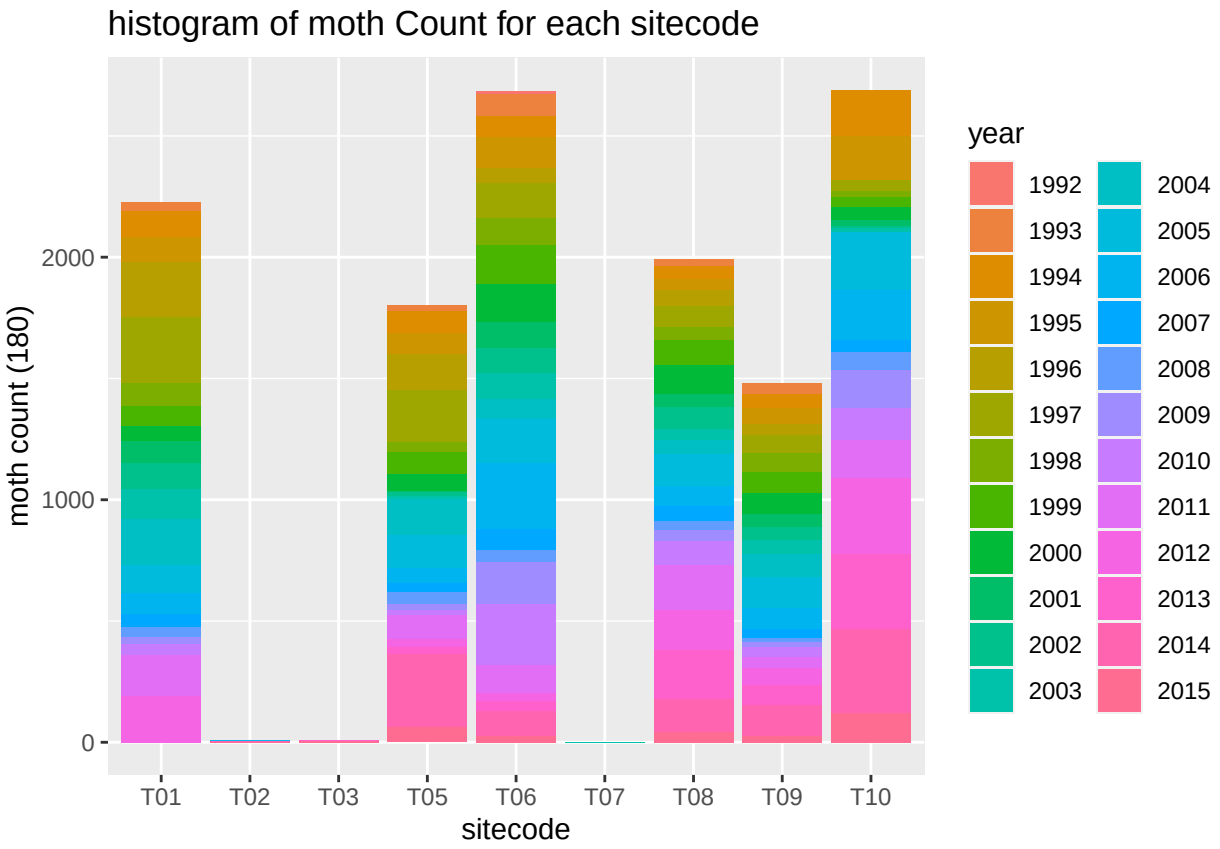
```
# for the "180" species we have moth count against year for each sitecode
ggplot(filtered_moth_data180, aes(x = factor(year), y = VALUE, fill = SITECODE)) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth count for each year",
       x = "Year", y = "moth count (180)", fill = "Sitecode")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```



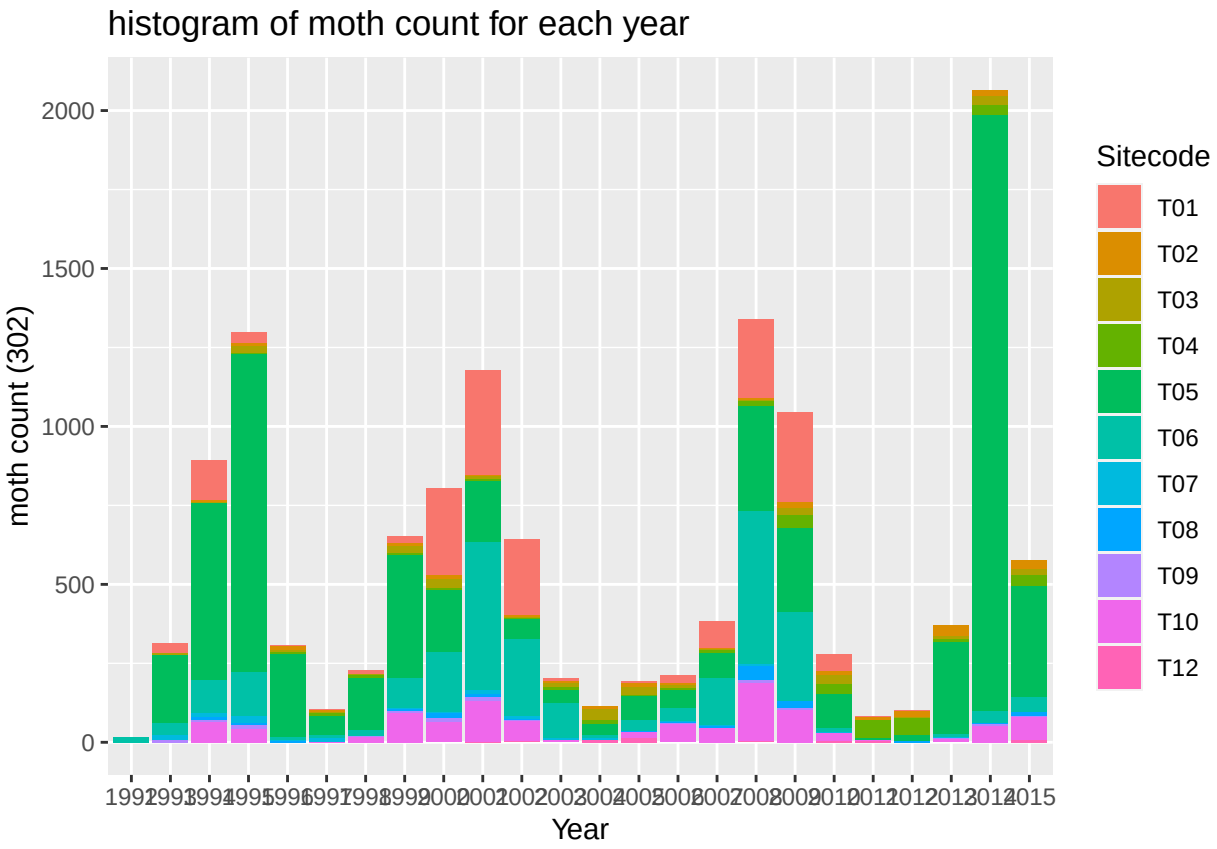

```
# moth count against sitecode for each year
ggplot(filtered_moth_data180, aes(x = SITECODE, y = VALUE, fill = factor(year))) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth Count for each sitecode",
       x = "sitecode", y = "moth count (180)", fill = "year")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```

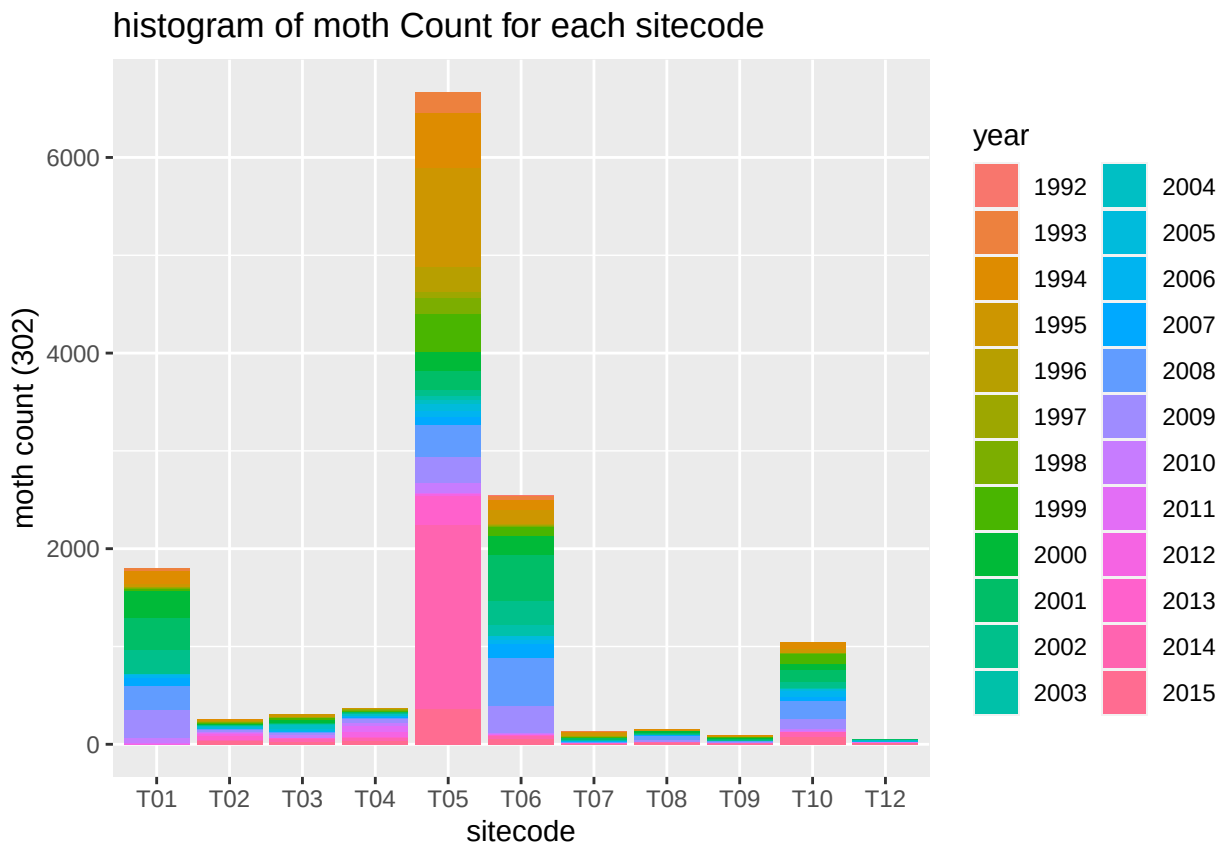



```
# same again for a different moth code so I will not comment on rest of similar code.
ggplot(filtered_moth_data302, aes(x = factor(year), y = VALUE, fill = SITECODE)) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth count for each year",
       x = "Year", y = "moth count (302)", fill = "Sitecode")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```

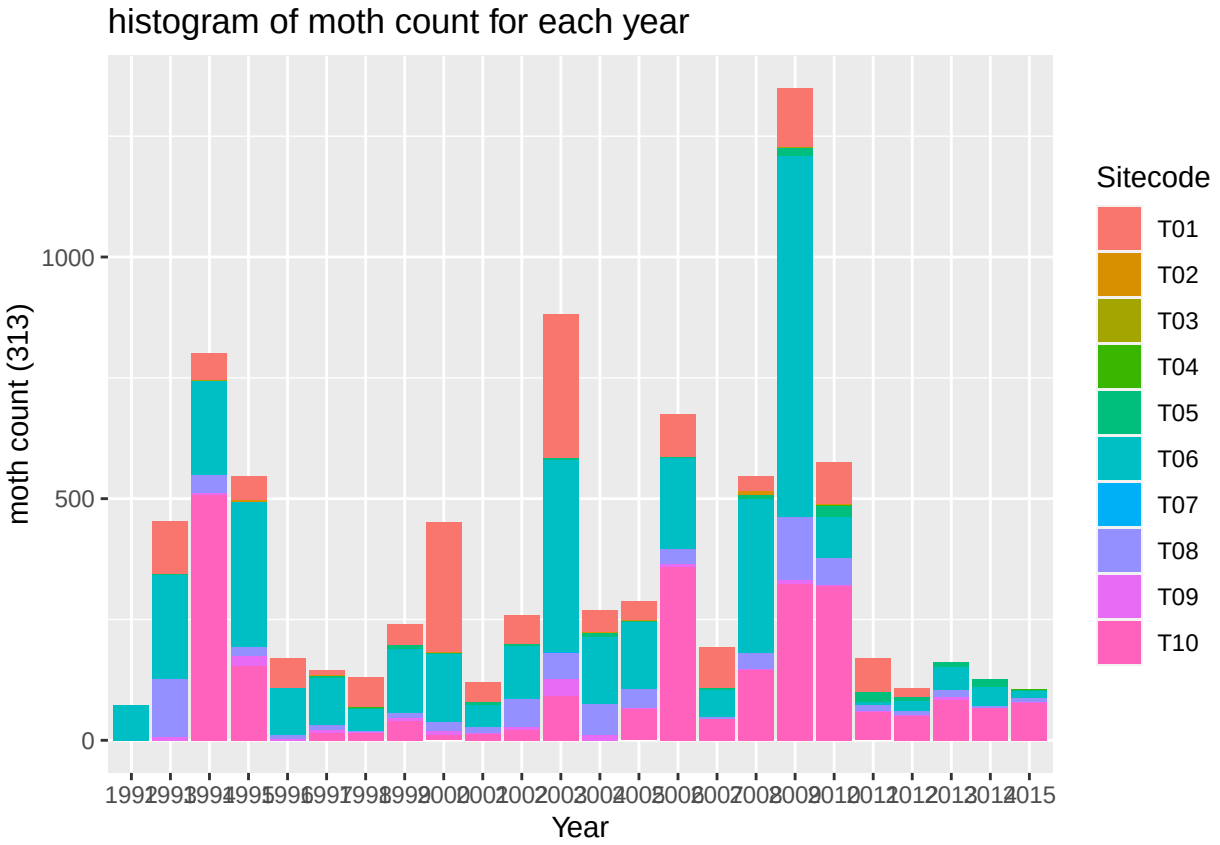



```
ggplot(filtered_moth_data302, aes(x = SITECODE, y = VALUE, fill = factor(year))) +
  geom_bar(stat = "identity", position = "stack") +
  labs(title = "histogram of moth Count for each sitecode",
       x = "sitecode", y = "moth count (302)", fill = "year")
```



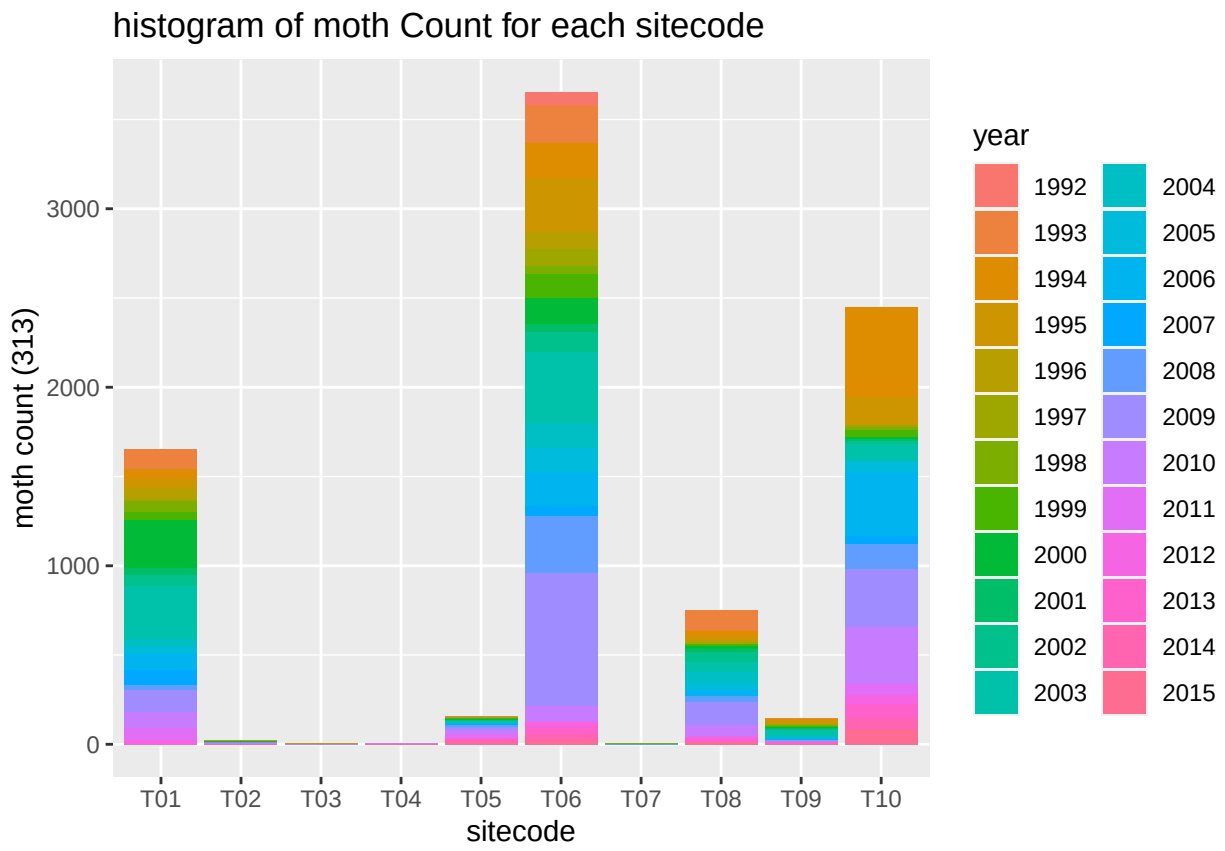

```
ggplot(filtered_moth_data313, aes(x = factor(year), y = VALUE, fill = SITECODE)) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth count for each year",
       x = "Year", y = "moth count (313)", fill = "Sitecode")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```



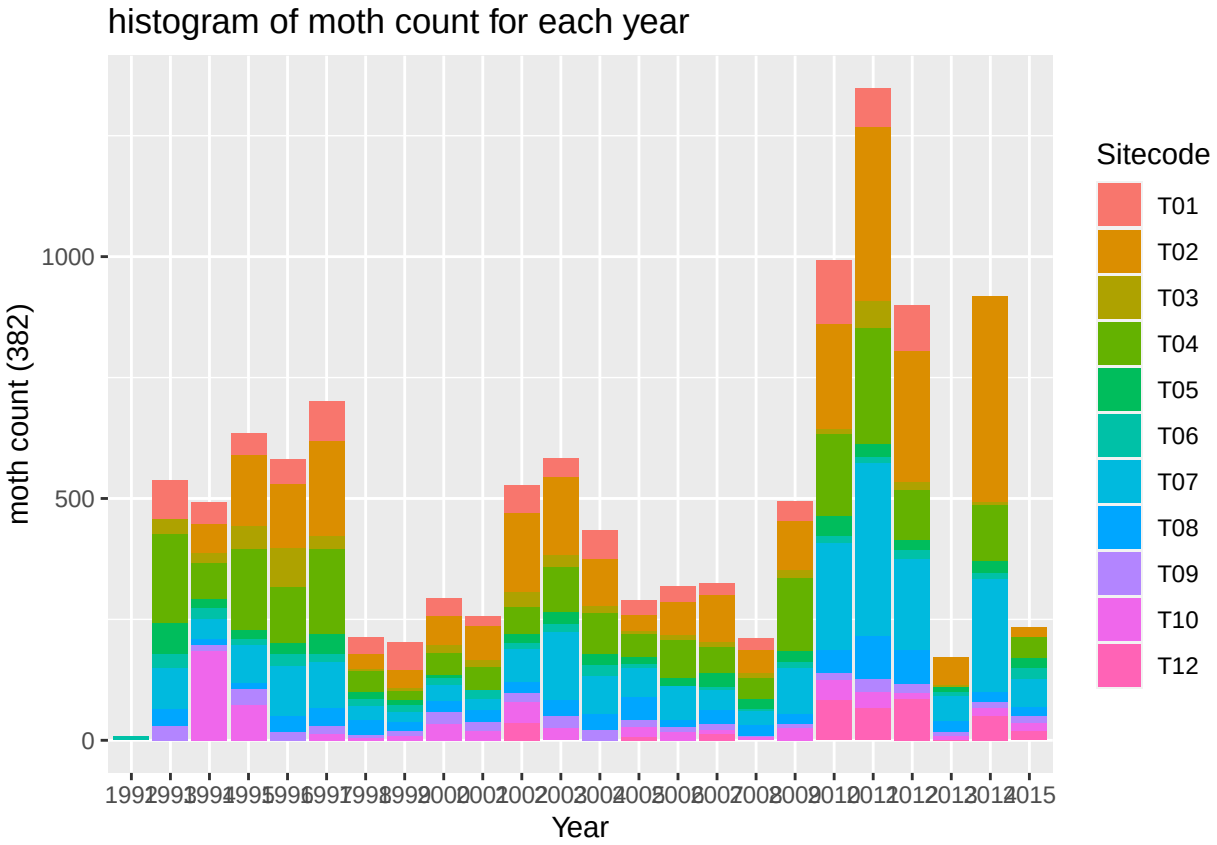

```
ggplot(filtered_moth_data313, aes(x = SITECODE, y = VALUE, fill = factor(year))) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth Count for each sitecode",
       x = "sitecode", y = "moth count (313)", fill = "year")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```



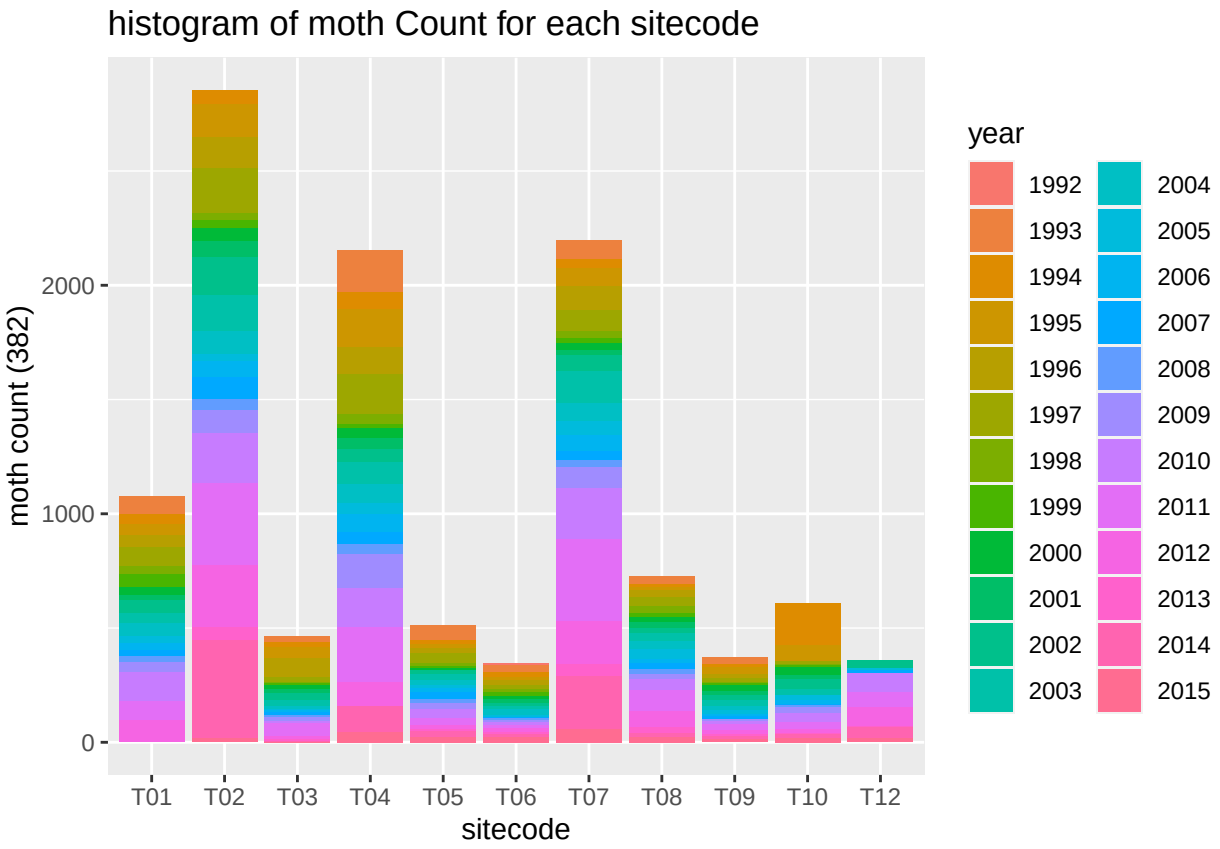

```
ggplot(filtered_moth_data382, aes(x = factor(year), y = VALUE, fill = SITECODE)) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth count for each year",
       x = "Year", y = "moth count (382)", fill = "Sitecode")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```



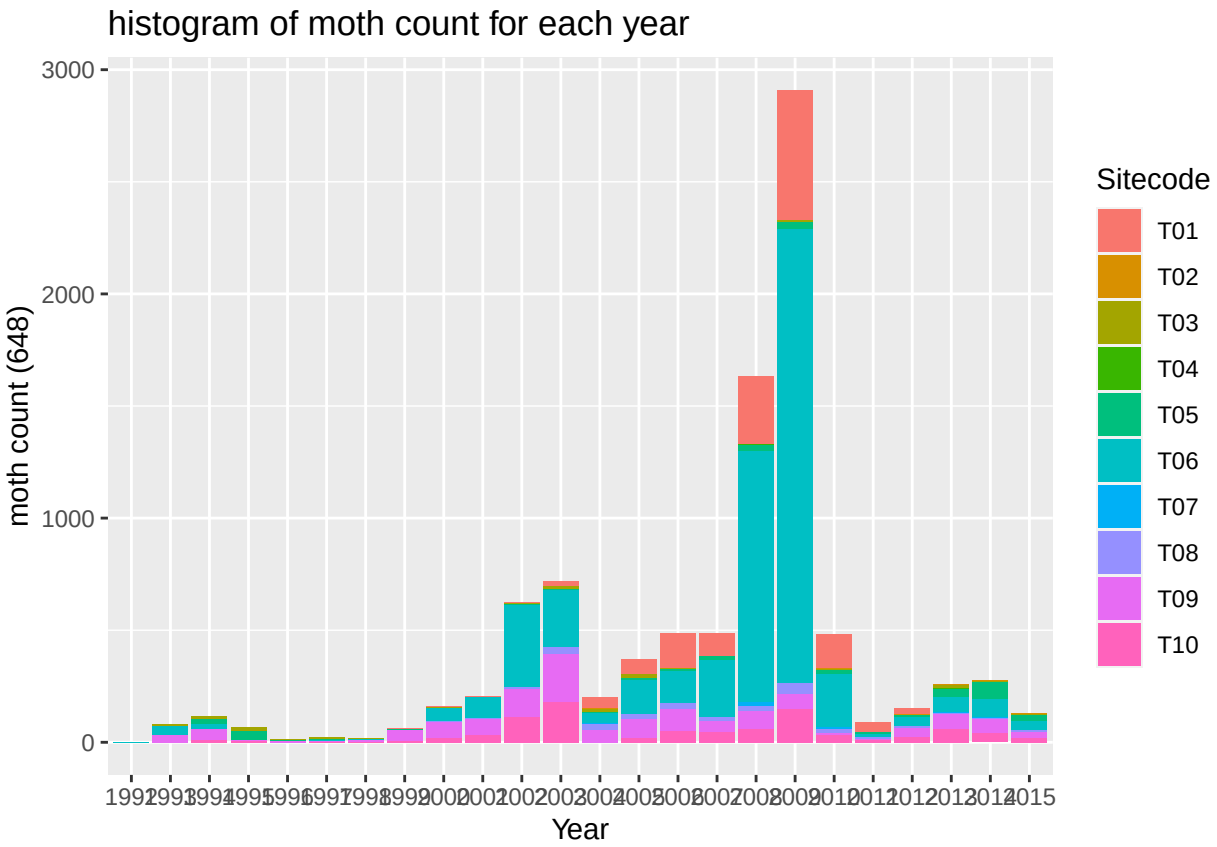

```
ggplot(filtered_moth_data382, aes(x = SITECODE, y = VALUE, fill = factor(year))) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth Count for each sitecode",
       x = "sitecode", y = "moth count (382)", fill = "year")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```



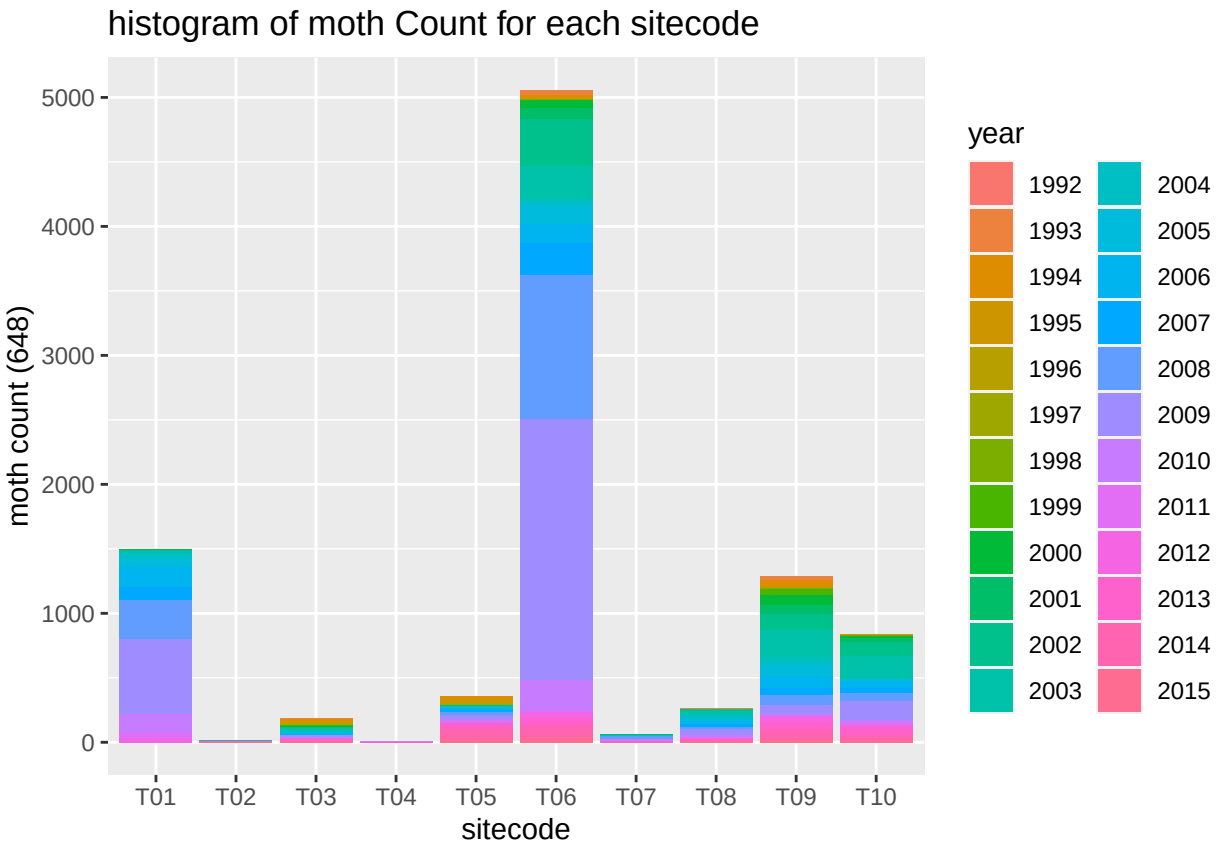

```
ggplot(filtered_moth_data648, aes(x = factor(year), y = VALUE, fill = SITECODE)) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth count for each year",
       x = "Year", y = "moth count (648)", fill = "Sitecode")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: `binwidth`, `bins`, and `pad`
```




```
ggplot(filtered_moth_data648, aes(x = SITECODE, y = VALUE, fill = factor(year))) +
  geom_histogram(stat = "identity", position = "stack") +
  labs(title = "histogram of moth Count for each sitecode",
       x = "sitecode", y = "moth count (648)", fill = "year")
```

```
## Warning in geom_histogram(stat = "identity", position = "stack"): Ignoring
## unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘
```



c)

Upper limit: 24 figures (e.g, for each of the 4 locations: 6 figures), 1 table, 200 words

```
# to find the most frequent moth species over all data
sum(filtered_moth_data180$VALUE)
```

```
## [1] 12888
```

```
sum(filtered_moth_data302$VALUE) # most frequent
```

```
## [1] 13394
```

```
sum(filtered_moth_data313$VALUE)
```

```
## [1] 8836
```

```
sum(filtered_moth_data382$VALUE)
```

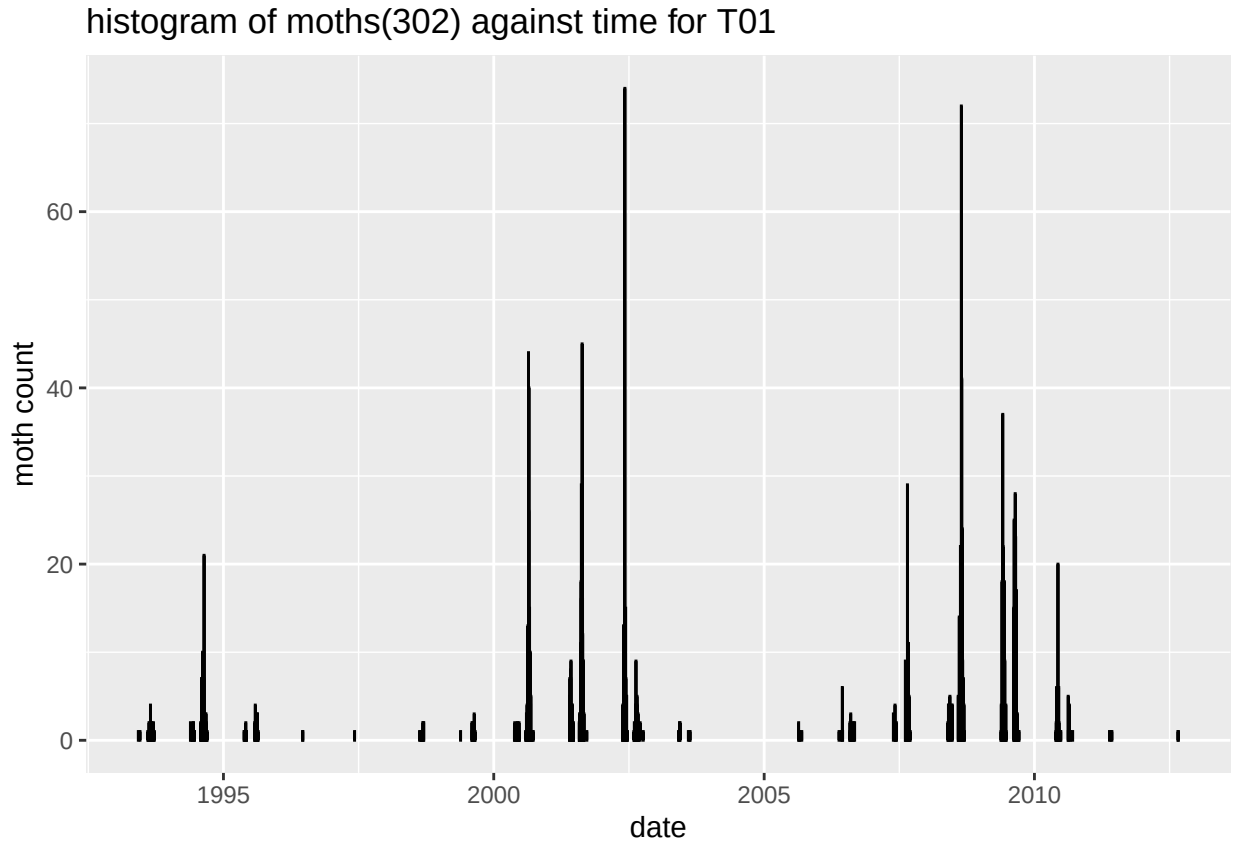
```
## [1] 11669
```

```r
sum(filtered_moth_data648$VALUE)
```

```
## [1] 9547
```

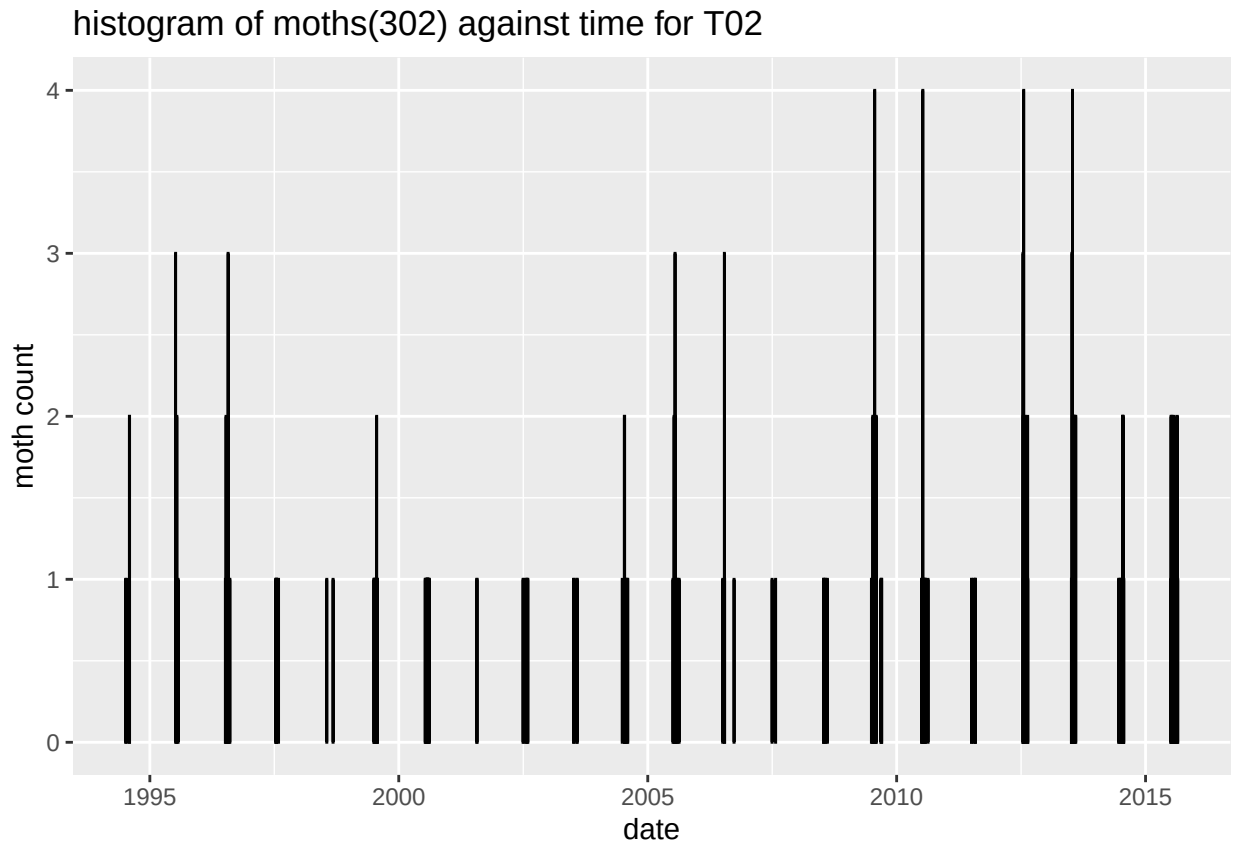
```r
# then i filtered the data for each of the sitecodes

T01_302_data <- filter(filtered_moth_data302, SITECODE == "T01")
T02_302_data <- filter(filtered_moth_data302, SITECODE == "T02")
T03_302_data <- filter(filtered_moth_data302, SITECODE == "T03")
T04_302_data <- filter(filtered_moth_data302, SITECODE == "T04")
T05_302_data <- filter(filtered_moth_data302, SITECODE == "T05")
T06_302_data <- filter(filtered_moth_data302, SITECODE == "T06")
T07_302_data <- filter(filtered_moth_data302, SITECODE == "T07")
T08_302_data <- filter(filtered_moth_data302, SITECODE == "T08")
T09_302_data <- filter(filtered_moth_data302, SITECODE == "T09")
T10_302_data <- filter(filtered_moth_data302, SITECODE == "T10")
T11_302_data <- filter(filtered_moth_data302, SITECODE == "T11")
T12_302_data <- filter(filtered_moth_data302, SITECODE == "T12")

# I will then use these new filtered data frames from the raw moth data frame to
# make histograms for each of the sitecodes

#T01- the codes are exactly the same for each histogram only the sitecodes are changes so I will not co

ggplot(T01_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T01", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘
```



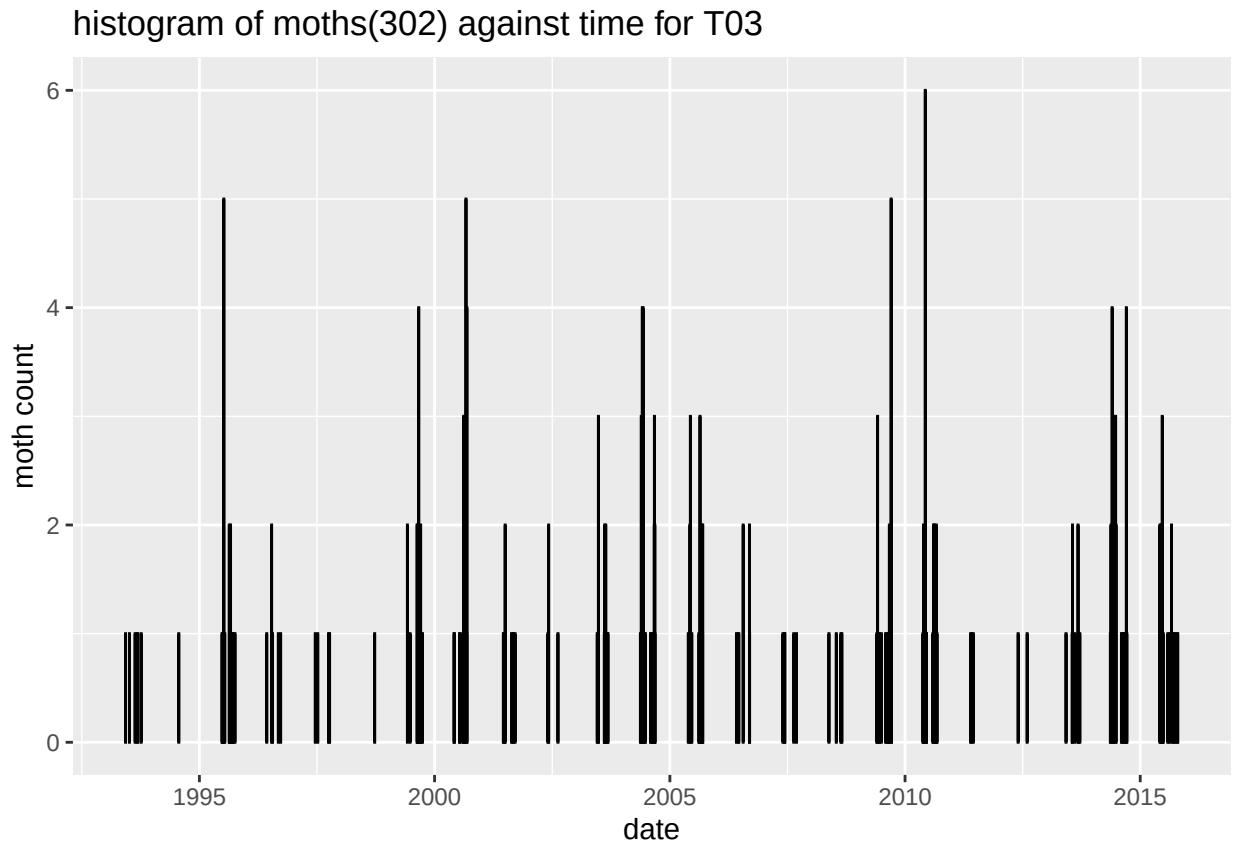

```
#T02
ggplot(T02_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T02", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



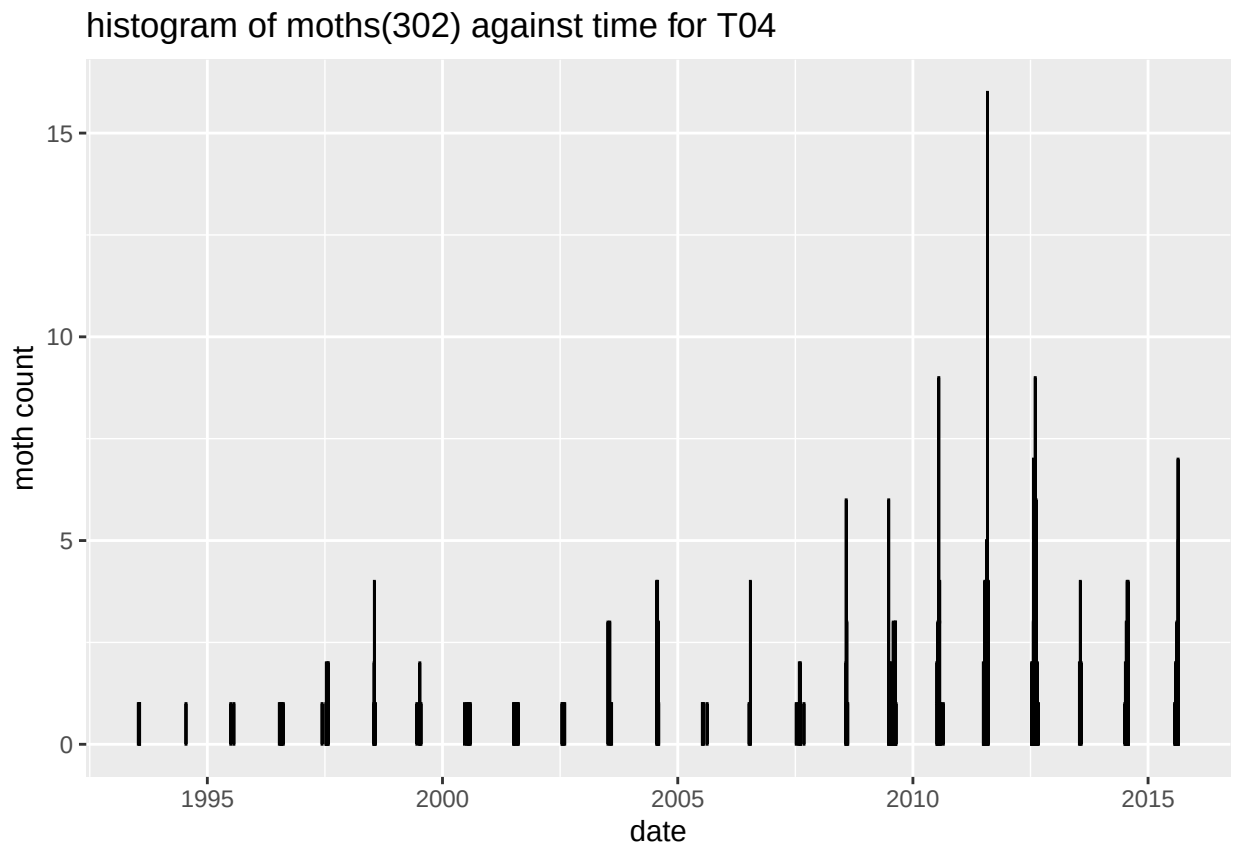

```
#T03
ggplot(T03_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T03", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



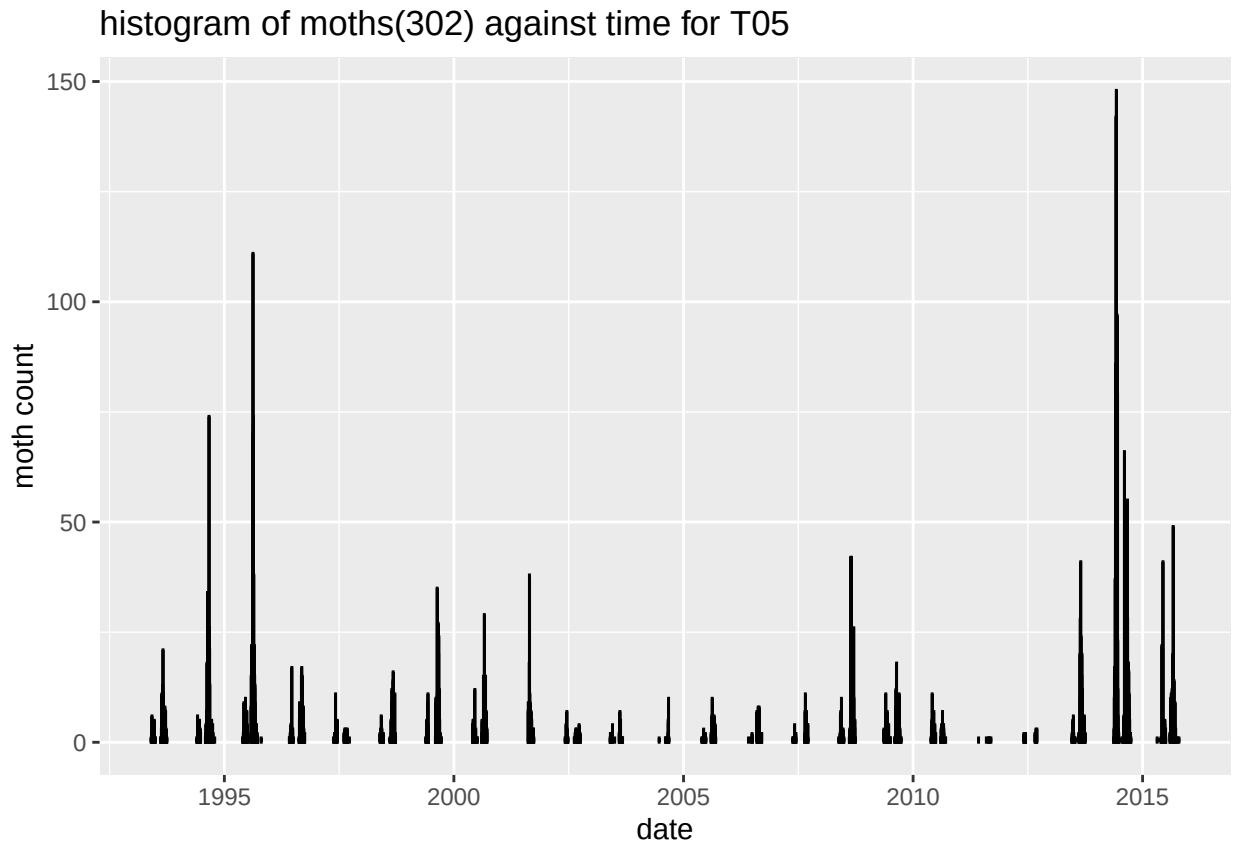

```
#T04
ggplot(T04_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T04", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



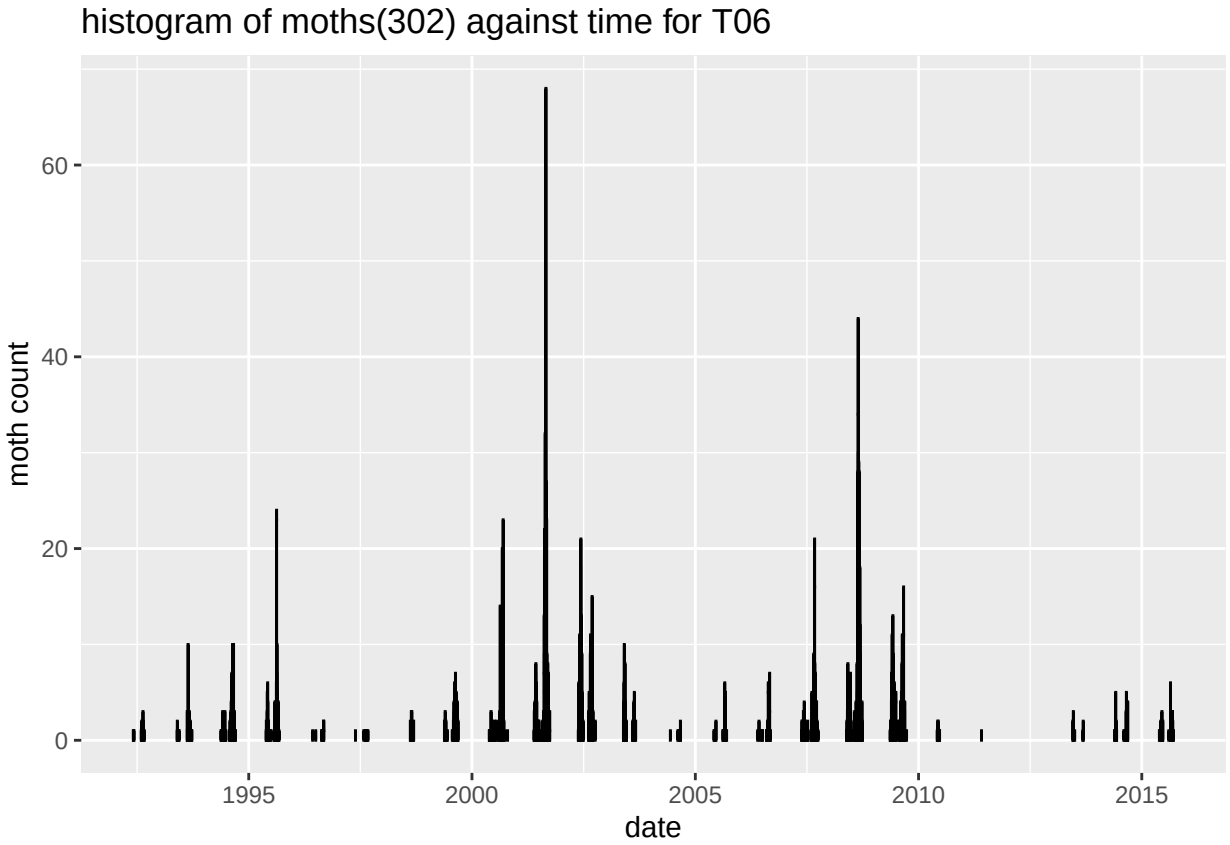

```
#T05
ggplot(T05_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T05", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



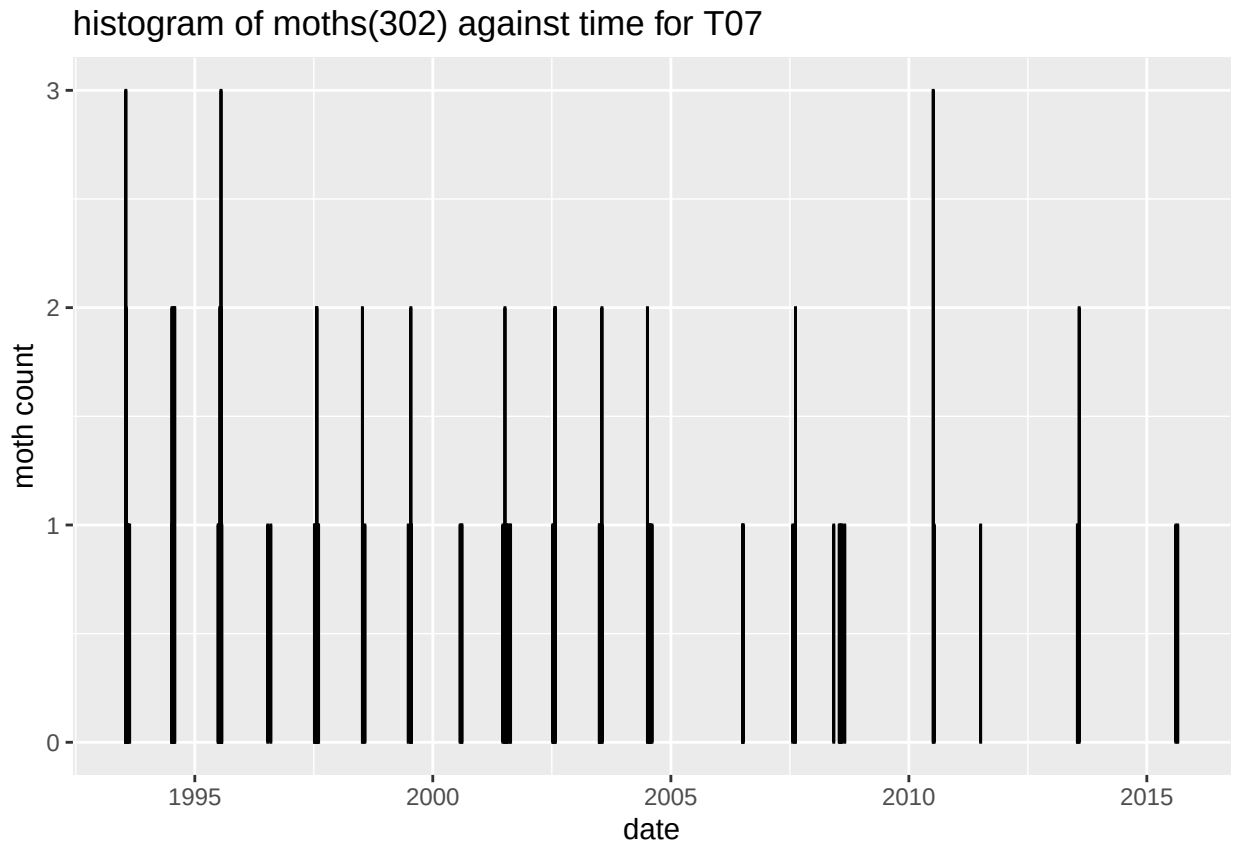

```
#T06
ggplot(T06_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T06", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



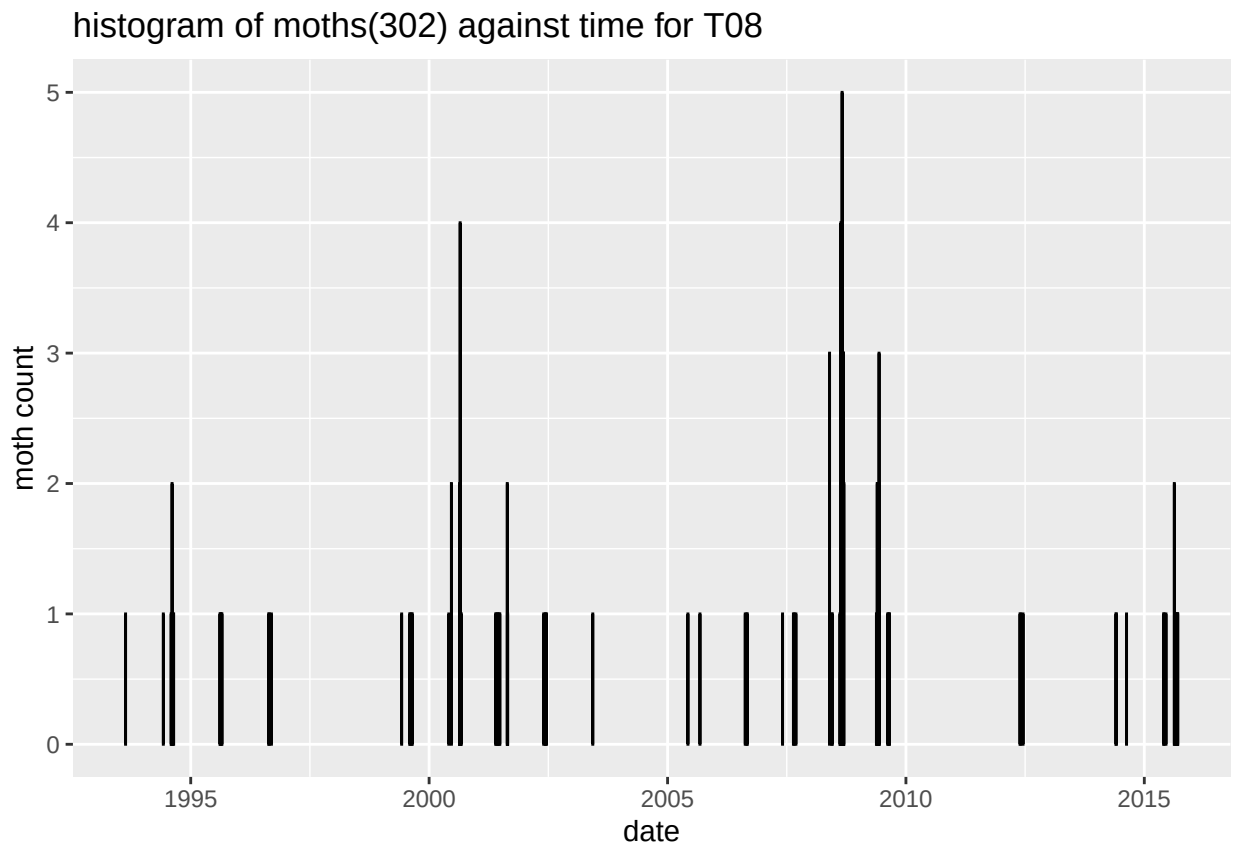

```
#T07
ggplot(T07_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T07", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



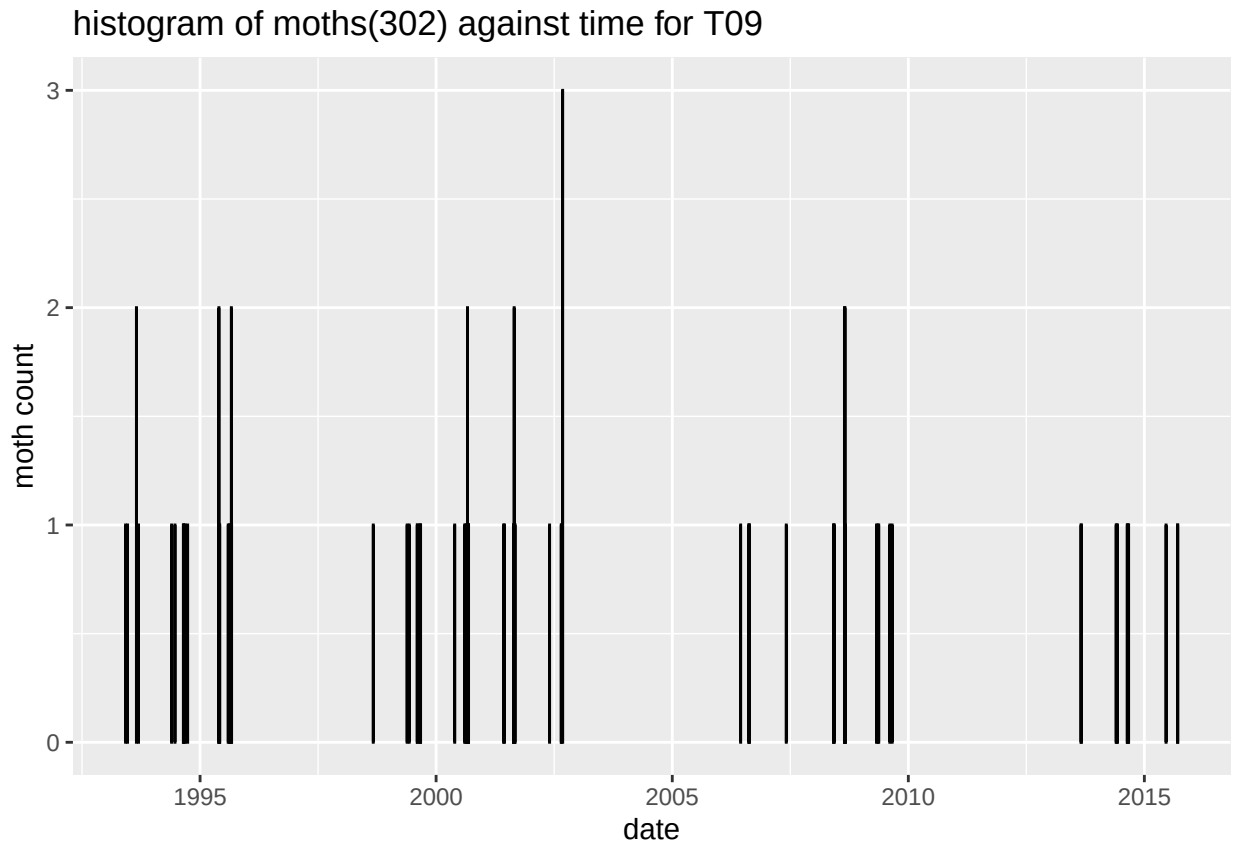

```
#T08
ggplot(T08_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T08", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```



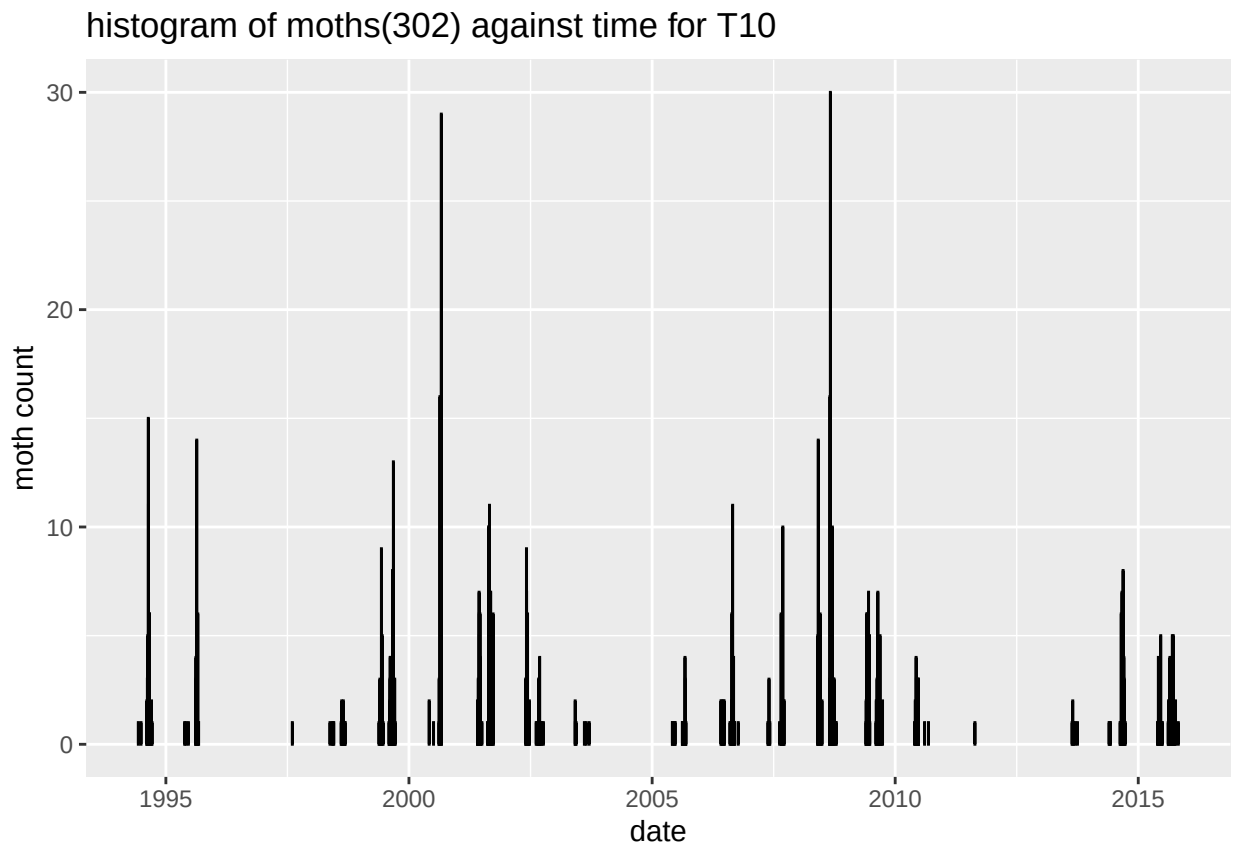

```
#T09
ggplot(T09_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T09", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```




```
#T10
ggplot(T10_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T10", x = "date", y = "moth count")
```

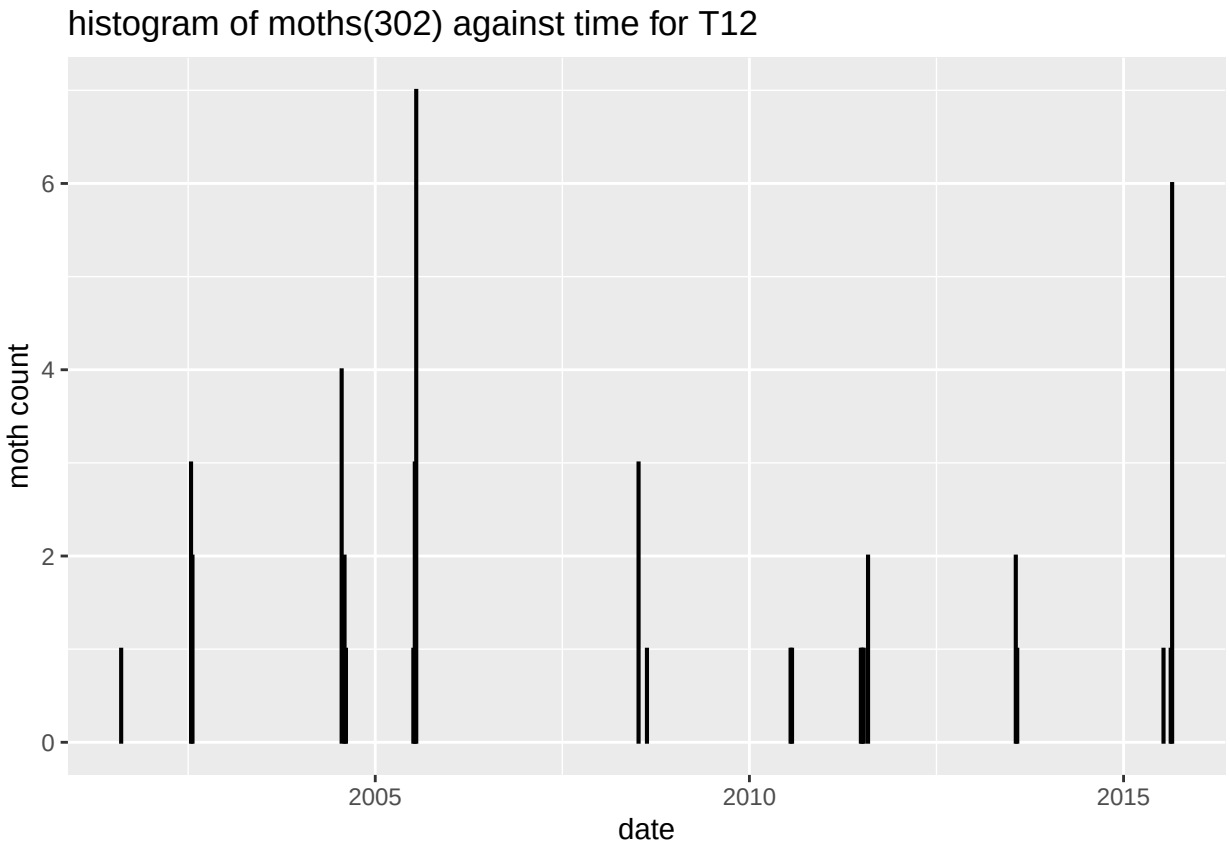
```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```




```
#T11
 #ggplot(T11_302_data, aes(x = date, y = VALUE)) +
 # geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
 # labs(title = "histogram of moths(302) against time for T11", x = "date", y = "moth count")
# there is no data for T11 for this species of moth over the whole time period
```

```
#T12
ggplot(T12_302_data, aes(x = date, y = VALUE)) +
  geom_histogram(stat = "identity", fill = "skyblue", color = "black") +
  labs(title = "histogram of moths(302) against time for T12", x = "date", y = "moth count")
```

```
## Warning in geom_histogram(stat = "identity", fill = "skyblue", color =
## "black"): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```


histogram of moths(302) against time for T12
moth count
0
2
4
6
2005
2010
2015
date