

ggplot2 Presentation - Computational Biology

P. Alexander Burnham

April 26, 2017

Preliminaries:

```
#####  
  
# Clear memory of variables and objects:  
rm(list=ls())  
  
# load packages:  
library(ggplot2)  
library(plyr)  
  
#####
```

Create a Data set:

```
# create sample IDs  
ID <- c(1:200)  
  
# create a factor called origin  
Origin <- c(rep("local", 100),  
            rep("California", 100))  
  
# create a factor called flower type  
FlowerType <- rep(c(rep("clover",25),  
                    rep("goldenrod",25),  
                    rep("trefoil",25),  
                    rep("mixed",25)),2)  
  
# create a variable called mass  
Mass <- c(rnorm(n = 100,  
              mean=32,  
              sd = 8), rnorm(n = 100,  
                            mean=21,  
                            sd=4))  
  
# create a variable called Nosema Load  
NosemaLoad <-c(rnorm(n = 100,  
                  mean=100000,  
                  sd = 80000), rnorm(n = 100,  
                                    mean=500000,  
                                    sd=40000))  
  
# create a variable called varroa load  
VarroaLoad <- c(rnorm(n = 100,  
                  mean=5,  
                  sd = 2), rnorm(n = 100,
```

```

                                mean=9,
                                sd=3))

# create a variable called time
Time <- rep(c(rep("Time1", 50), rep("Time2", 50)),2)

# create data frame
DF <- data.frame(ID, Origin, FlowerType, Mass, NosemaLoad, VarroaLoad, Time)

```

Basic Graphics in ggplot:

Summary stats using ddply:

```

# using ddply to get summary stats for mass:
DF1 <- ddply(DF, c("FlowerType"), summarise,
             n = length(Mass),
             mean = mean(Mass, na.rm=TRUE),
             sd = sd(Mass, na.rm=TRUE),
             se = sd / sqrt(n))

```

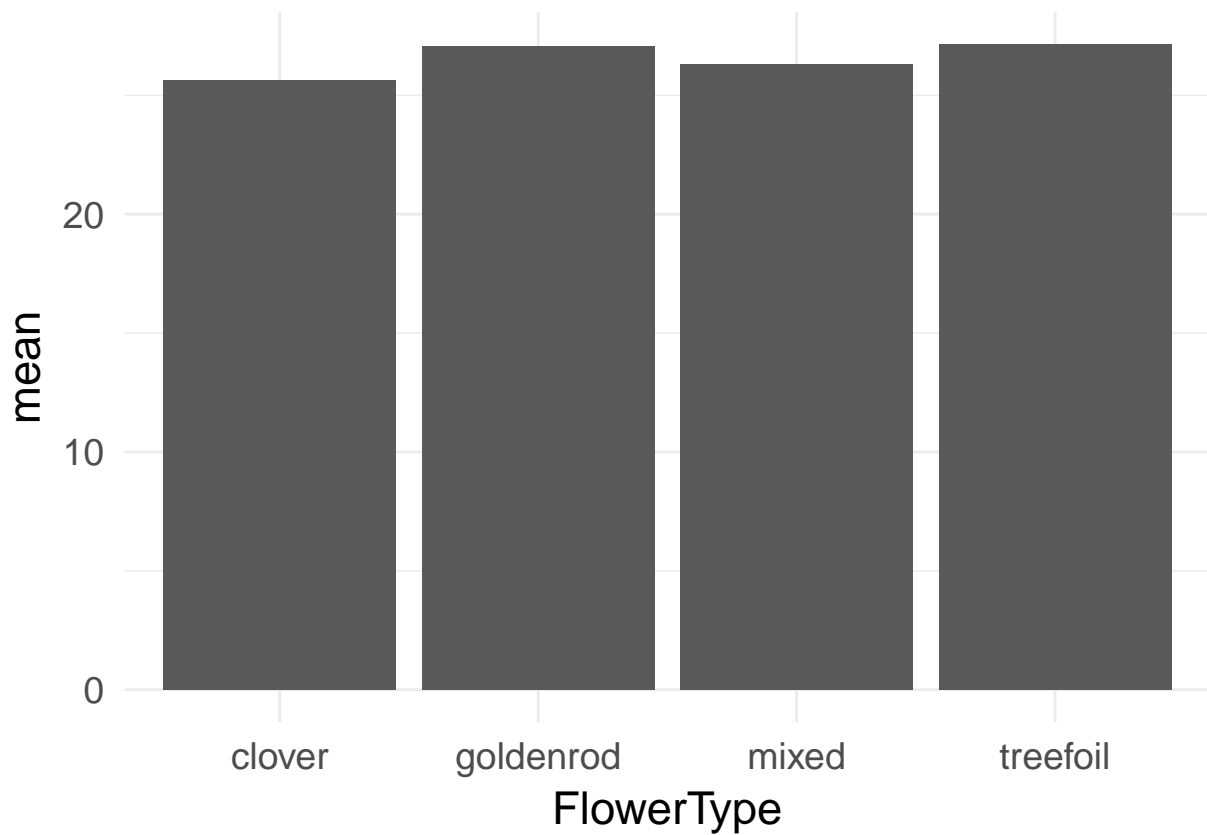
Bar Plot

```

plot1 <- ggplot(DF1, aes(x=FlowerType,
                        y=mean)) + geom_bar(stat = "identity")

plot1 + theme_minimal(base_size = 17)

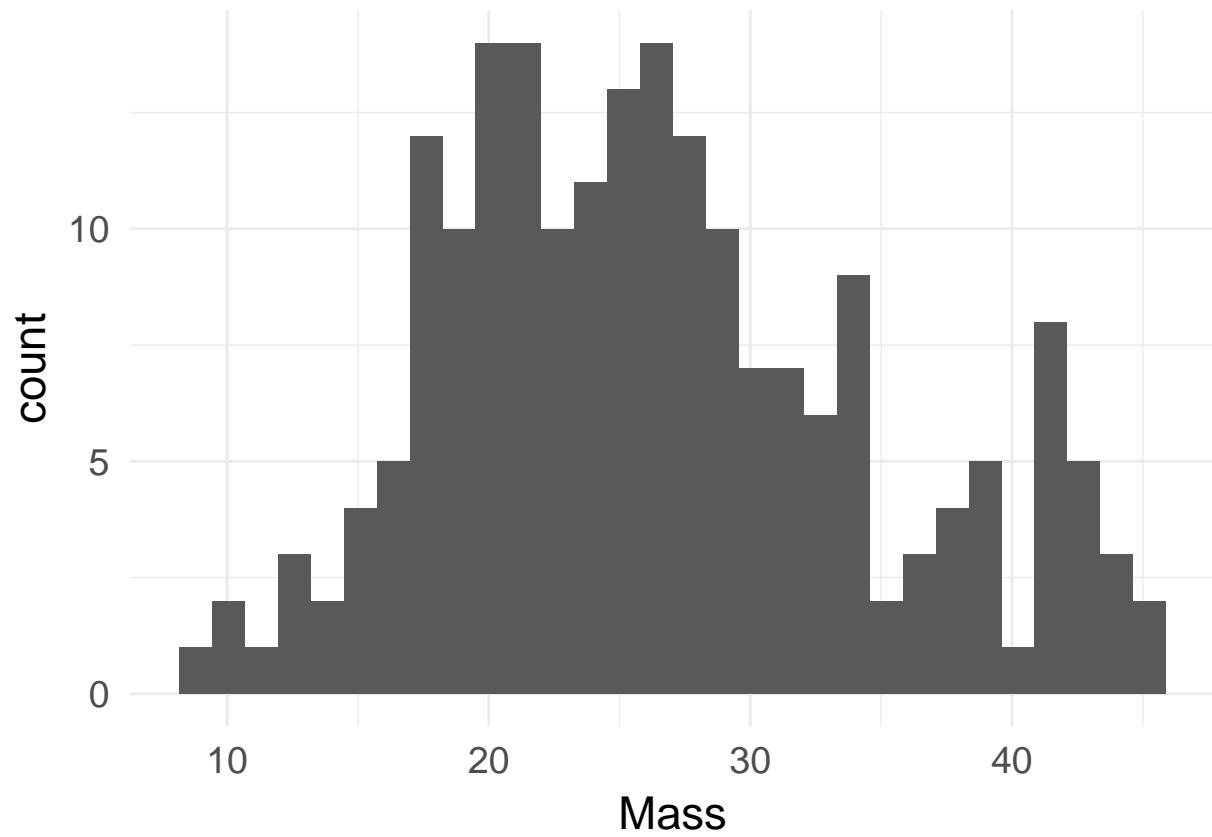
```



```
# + coord_cartesian(ylim = c(0, 40))
```

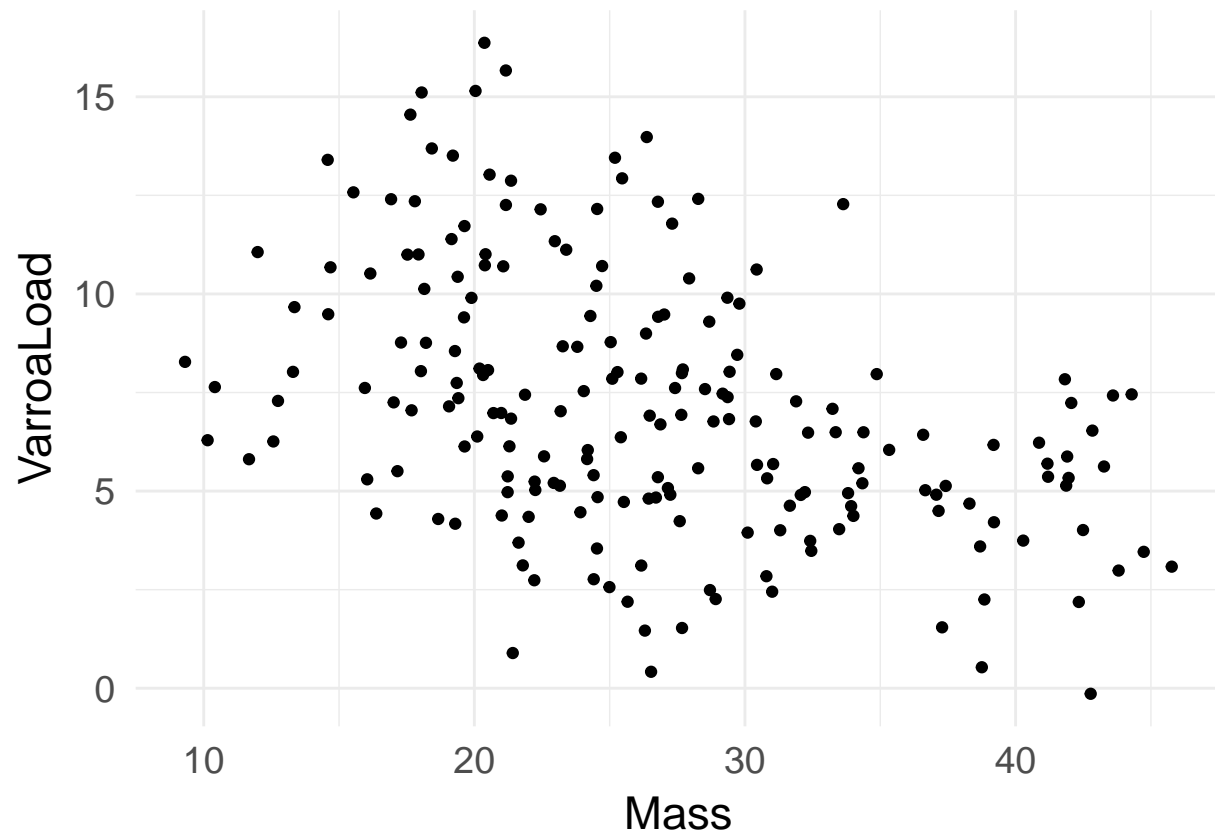
Histogram

```
plot2 <- ggplot(DF, aes(Mass))  
plot2 + geom_histogram() + stat_bin(bins = 30) + theme_minimal(base_size = 17)  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



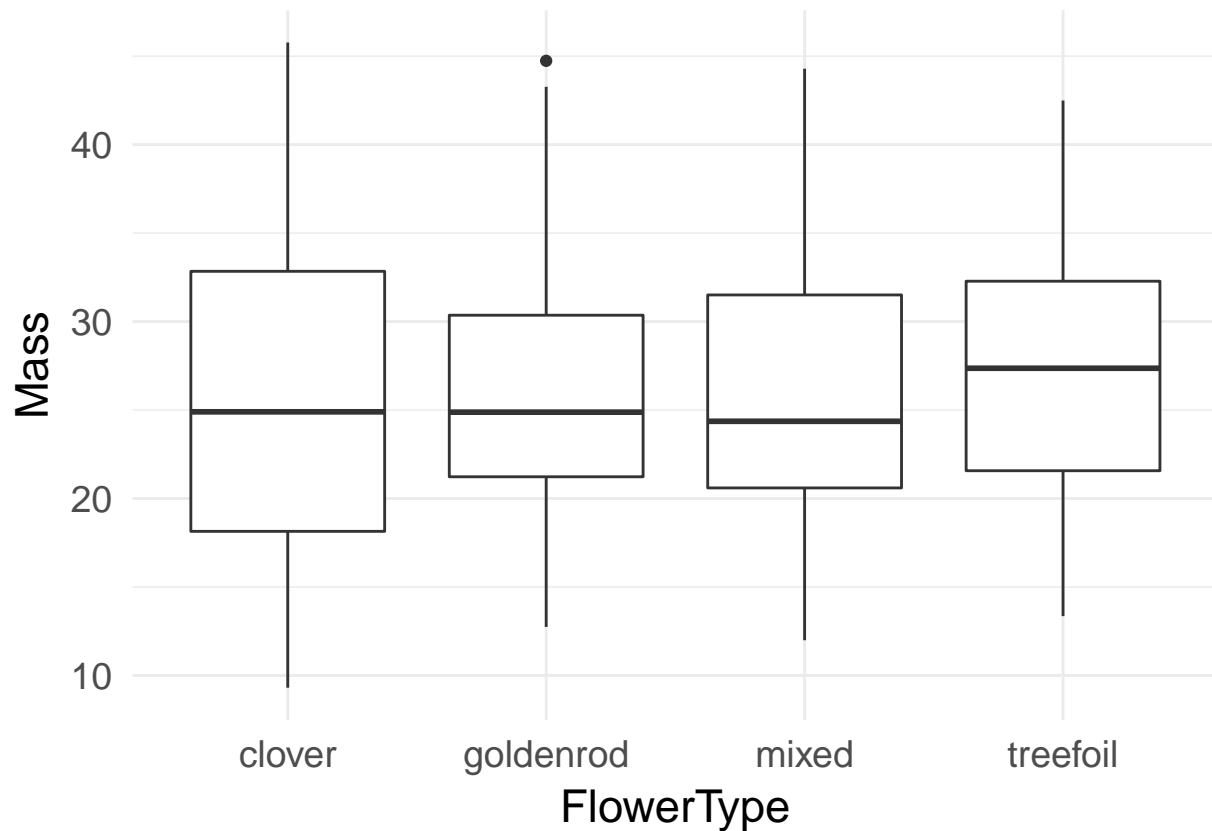
Scatterplot

```
plot3 <- ggplot(DF, aes(x=Mass,  
                        y=VarroaLoad))  
  
plot3 + geom_point() + theme_minimal(base_size = 17)
```



Boxplot

```
plot4 <- ggplot(DF, aes(x=FlowerType,  
                        y=Mass))  
  
plot4 + geom_boxplot() + theme_minimal(base_size = 17)
```



More Complicated Graphics in ggplot

Summary stats using ddply:

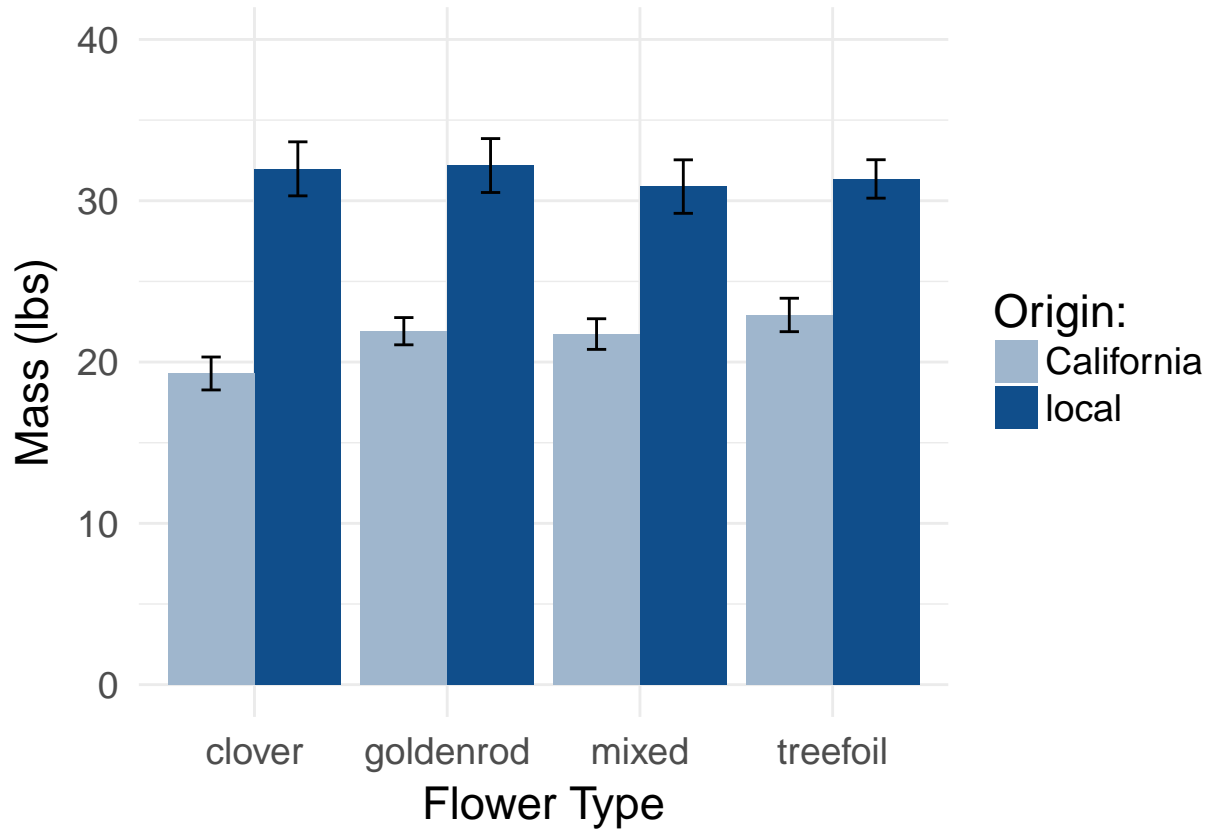
```
# using ddply to get summary stats for mass:
DF2 <- ddply(DF, c("FlowerType", "Origin"), summarise,
  n = length(Mass),
  mean = mean(Mass, na.rm=TRUE),
  sd = sd(Mass, na.rm=TRUE),
  se = sd / sqrt(n))
```

Bar Plot

```
#choosing color pallet
colors <- c("slategray3", "dodgerblue4")

#Create a bar graph for with CI and SE bars
plot5 <- ggplot(DF2, aes(x=FlowerType,
  y=mean,
  fill=Origin)) +
  geom_bar(stat="identity",
    position=position_dodge()) + labs(x="Flower Type", y = "Mass (lbs)") + geom_errorbar(aes(ymin=
```

```
plot5 + theme_minimal(base_size = 17) + scale_fill_manual(values=colors, name="Origin:") + coord_cartes
```

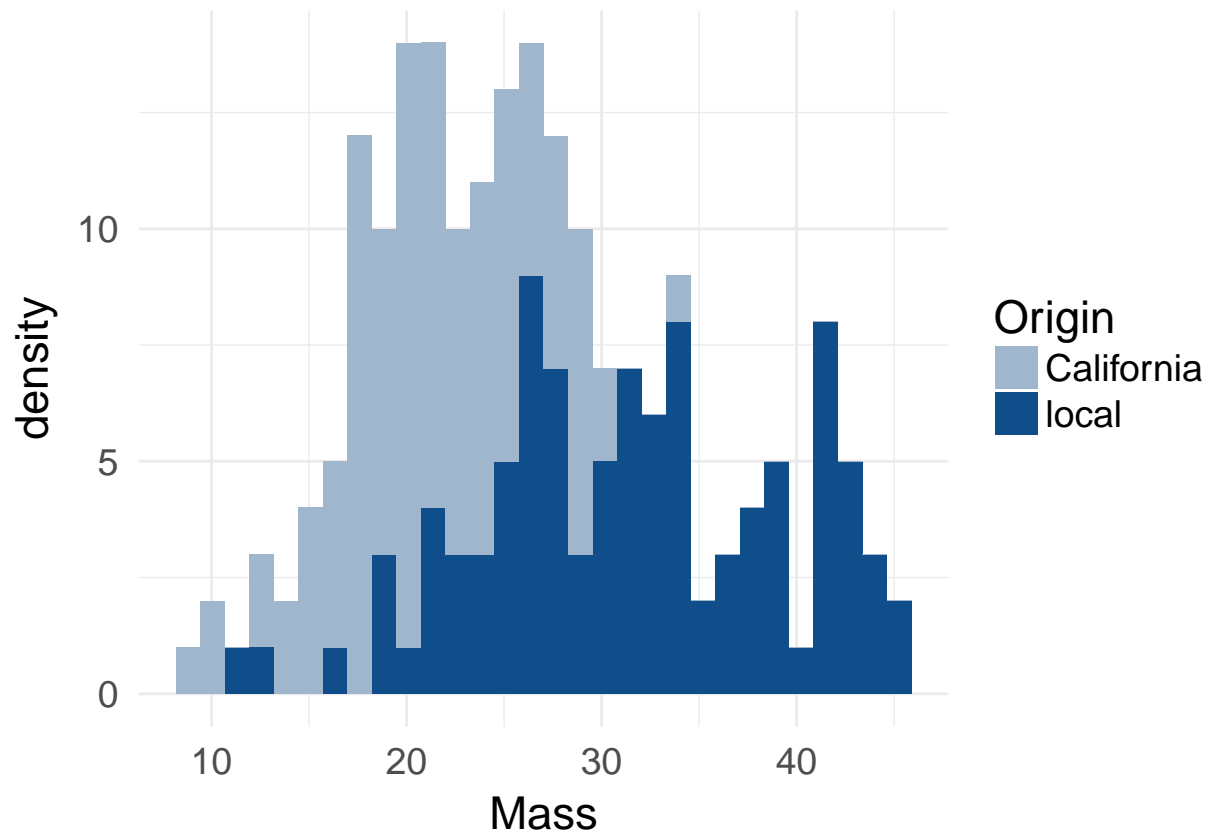


Histogram

```
plot6 <- ggplot(DF, aes(Mass, fill=Origin)) +
  geom_histogram(alpha=0.2, aes(y = ..density..), position="identity")

plot6 + stat_bin(bins = 30) + theme_minimal(base_size = 17) + scale_fill_manual(values=colors)

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

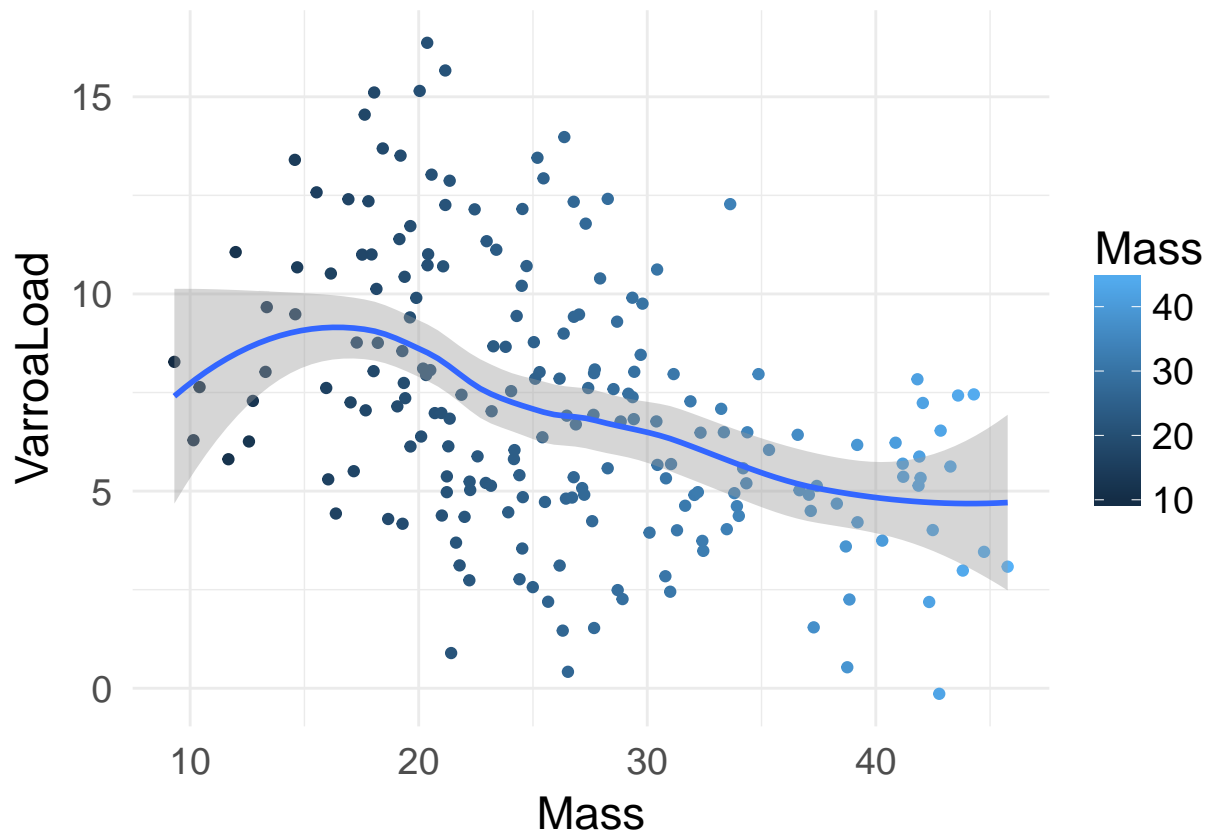


Scatterplot

```
plot7 <- ggplot(DF, aes(x=Mass,
                        y=VarroaLoad))

plot7 + geom_point(aes(color = Mass)) + theme_minimal(base_size = 17) + geom_smooth()

## `geom_smooth()` using method = 'loess'
```

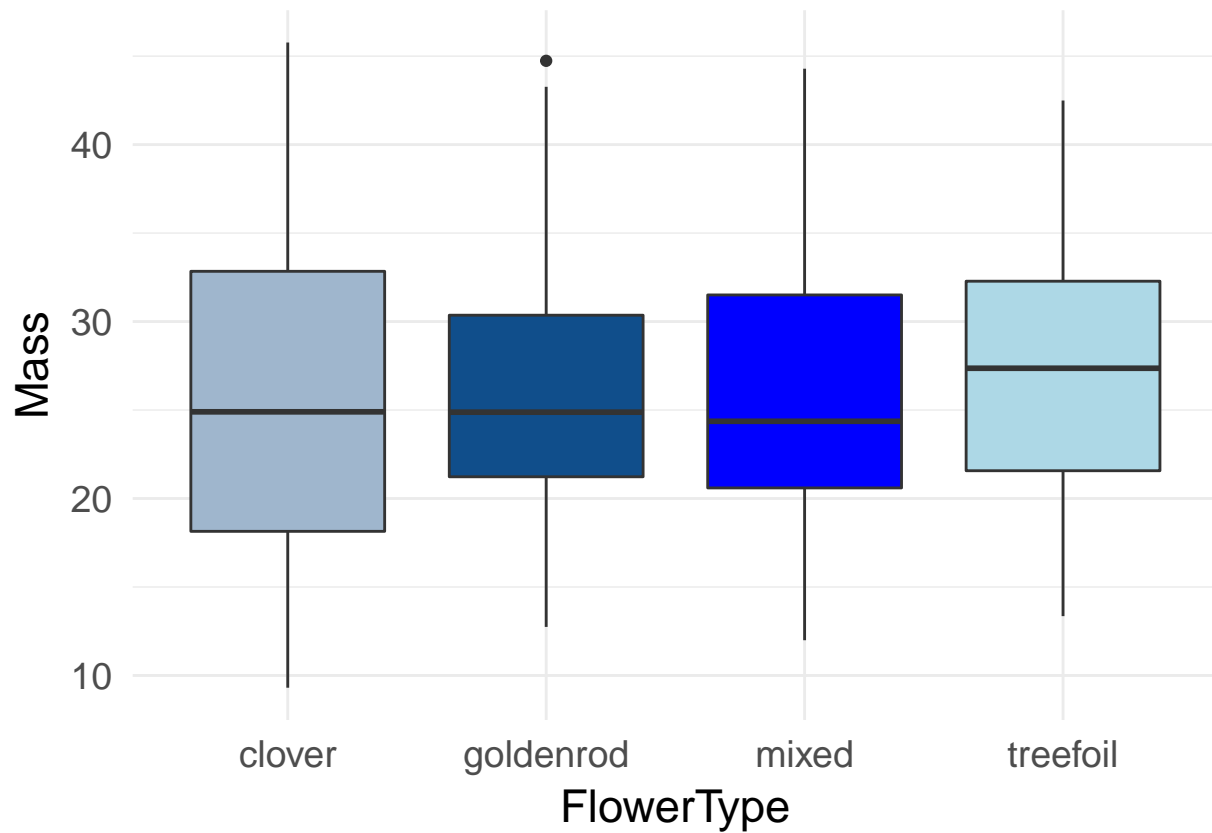
```
# method = "lm", se = TRUE
```

Boxplot

```
colors1 <- c("slategray3", "dodgerblue4", "blue", "lightblue")

plot8 <- ggplot(DF, aes(x=FlowerType,
                        y=Mass,
                        fill=FlowerType))

plot8 + geom_boxplot() + scale_fill_manual(values=colors1, guide_legend(NULL)) + guides(fill=FALSE) + t
```



Clean graphics for publication:

```
DF3 <- ddply(DF, c("Time", "Origin"), summarise,
  n = length(NosemaLoad),
  mean = mean(NosemaLoad, na.rm=TRUE),
  sd = sd(NosemaLoad, na.rm=TRUE),
  se = sd / sqrt(n))

plot9 <- ggplot(data = DF3,
  aes(x = Time,
    y = mean,
    group = Origin)
) + geom_point(size=3) + scale_colour_manual(values = c("black", "black")) + labs(x = "Time", y = "Nosema Load")
plot9 + theme(legend.position=c(.15, .85)) + labs(linetype="Queen Origin")
```

