Chickweight Plotting Exercise

Madeline Cowen

September 15, 2016

I. Use the ChickWeight R dataset to examine if there is an effect of diet on chick growth:

```
chickweight=ChickWeight
chickweight <- within(chickweight,{
  Chick <- factor(Chick)
  #Time <- factor(Time)
  Diet <- factor(Diet)
})
```

1. Plot and examine if there is a statistical difference between the weights of chicks assigned to different diet treatments on day 0.
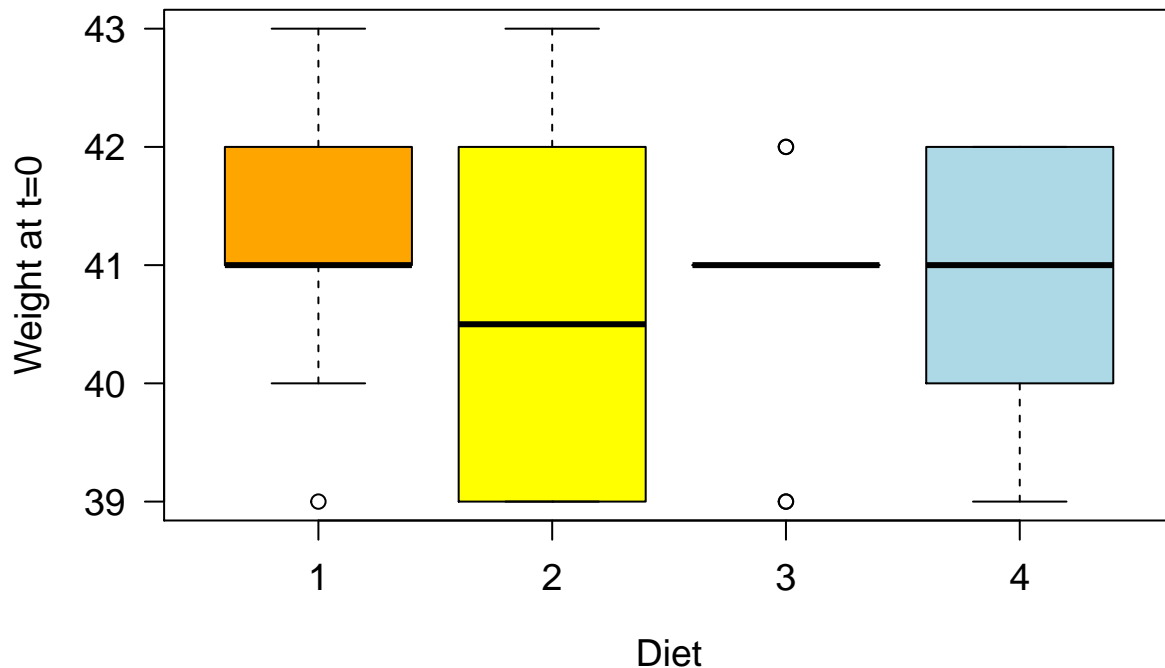
```
day0df <- chickweight[chickweight$Time==0,]
summary(day0df) #there are four diets
```

```
##      weight          Time         Chick      Diet
##  Min.   :39.00   Min.   :0    18     : 1    1:20
##  1st Qu.:41.00   1st Qu.:0    16     : 1    2:10
##  Median :41.00   Median :0    15     : 1    3:10
##  Mean   :41.06   Mean   :0    13     : 1    4:10
##  3rd Qu.:42.00   3rd Qu.:0    9      : 1
##  Max.   :43.00   Max.   :0    20     : 1
##                               (Other):44
```

```
summary(anova1 <- aov(day0df$weight~day0df$Diet))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## day0df$Diet  3   4.32   1.440   1.132  0.346
## Residuals   46  58.50   1.272
```

```
plot(day0df$Diet, day0df$weight, # data to plot
    ylab="Weight at t=0",xlab="Diet", # label axes
      col=c("orange","yellow","pale green","light blue"),las=1, cex.axis=1.2,cex.lab=1.2)
```

2. Plot and examine if there is a statistical difference between the weights of chicks assigned to different diet treatments at the end of the study (day 21).

```
day21df <- chickweight[chickweight$Time==21,]
summary(day21df) #there are four diets
```

```
##      weight            Time          Chick     Diet
##  Min.   : 74.0   Min.   :21   13     : 1   1:16
##  1st Qu.:167.0   1st Qu.:21   9      : 1   2:10
##  Median :205.0   Median :21   20     : 1   3:10
##  Mean   :218.7   Mean   :21   10     : 1   4: 9
##  3rd Qu.:266.0   3rd Qu.:21   17     : 1
##  Max.   :373.0   Max.   :21   19     : 1
##                               (Other):39
```

```
summary(anova1 <- aov(day21df$weight~day21df$Diet))
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## day21df$Diet   3  57164   19055   4.655 0.00686 **
## Residuals     41 167839    4094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(day21df$Diet, day21df$weight, # data to plot
    ylab="Weight at t=21",xlab="Diet", # label axes
     col=c("orange","yellow","pale green","light blue"),las=1, cex.axis=1.2,cex.lab=1.2)
```
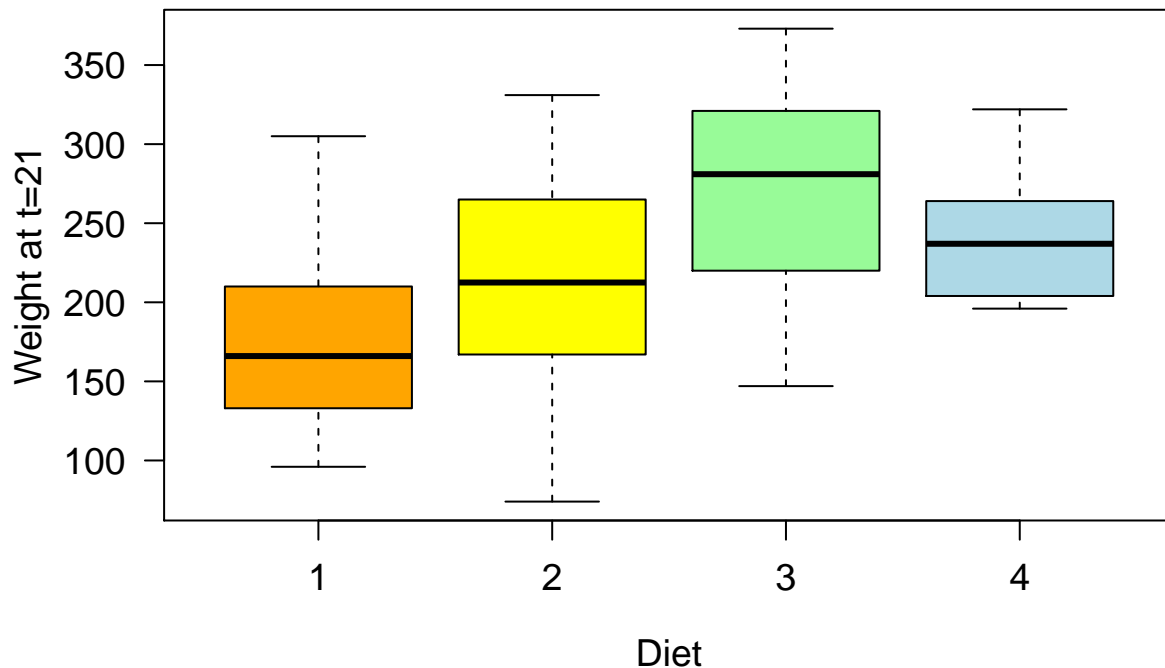
```
TukeyHSD(anova1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = day21df$weight ~ day21df$Diet)
##
## $`day21df$Diet`
##          diff        lwr       upr     p adj
## 2-1  36.95000  -32.11064 106.01064 0.4868095
## 3-1  92.55000   23.48936 161.61064 0.0046959
## 4-1  60.80556  -10.57710 132.18821 0.1192661
## 3-2  55.60000  -21.01591 132.21591 0.2263918
## 4-2  23.85556  -54.85981 102.57092 0.8486781
## 4-3 -31.74444 -110.45981  46.97092 0.7036249
```

3. Is there an effect of diet on chick growth? (HINT: you will need to use a repeated measures ANOVA)

```
summary(aov(weight~Diet*Time+Error(Chick),data=chickweight))
```

```
##
## Error: Chick
##            Df Sum Sq Mean Sq F value   Pr(>F)
## Diet        3 155863   51954   7.221 0.000482 ***
## Time        1  55175   55175   7.668 0.008197 **
## Diet:Time   1   2482    2482   0.345 0.559984
## Residuals  44 316586    7195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
```

```
##             Df  Sum Sq Mean Sq F value Pr(>F)
## Time         1 1962914 1962914 3049.28 <2e-16 ***
## Diet:Time    3   84222   28074   43.61 <2e-16 ***
## Residuals  524  337315     644
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. Plot the effects of diet on chick growth.
   Instead of a simple answer to this question, I present to you a story of struggling to find a more elegant solution.

```r
#First, I made separate plots for each diet.
par(mfrow=c(2,2))
chickcols <- c("orange","yellow","pale green","light blue")
for (i in 1:4){
  dietdf <- chickweight[chickweight$Diet==i,]
  plot(dietdf$Time,dietdf$weight,ylim=c(0,400),col=chickcols[i],xlab="Time (in days)",ylab="Weight")
text(8,300,paste('Diet', i),cex=1.5)
}
```
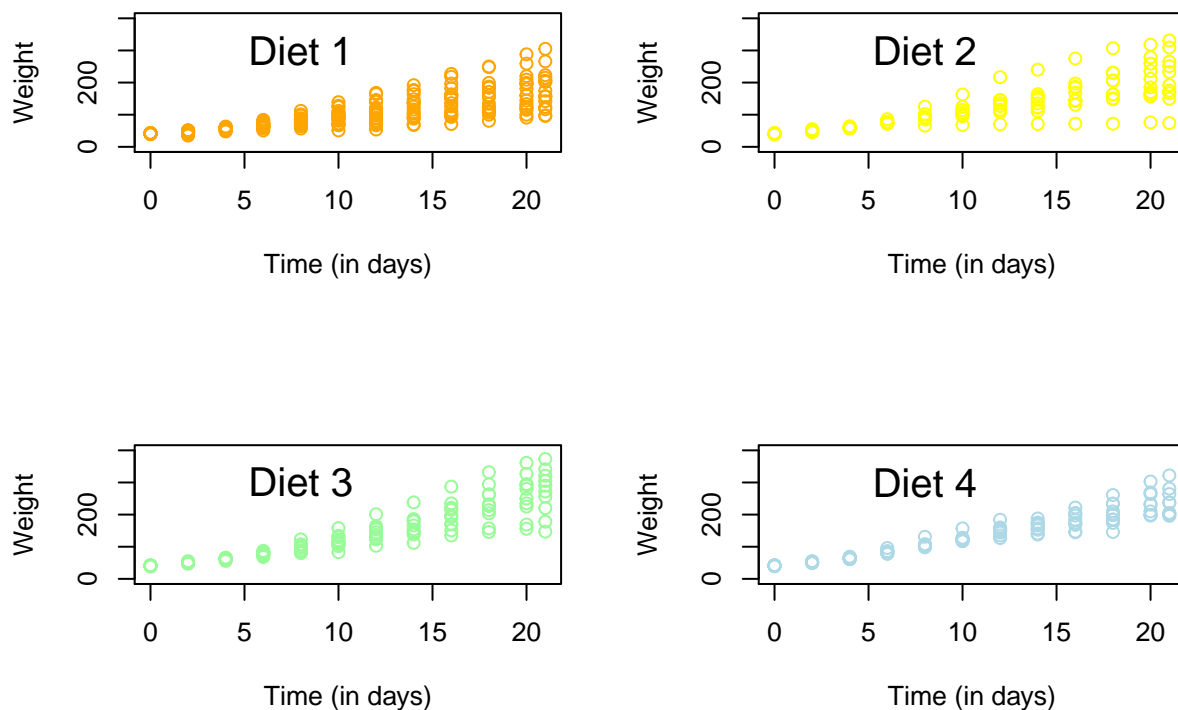


```r
#Even though they have the same y limits, they are not easy to compare, so I wanted to put them on the
par(mfrow=c(1,1))
timeWeights <- NA
avWeights <- NA
plot(1, type = "n", xlim = c(0, 23), ylim = c(0, 280),
     xlab = "Day", ylab = "Weight")   #create a blank plot that is the size I want
for (dietTypes in unique(chickweight$Diet)) {
  dietdf <- subset(chickweight,Diet==dietTypes)
  mean_weight_per_day <- numeric(max(chickweight$Time)+1)
  for (day in 0:max(chickweight$Time)){
```

4

```
        weights_for_this_day <- dietdf[dietdf$Time == day, 1] #take the first column (weights) from the die
        mean_weight_per_day[day+1] <- mean(weights_for_this_day, na.rm = T) #take the average of those weig
    }
    mean_weight_per_day <- mean_weight_per_day[which(!is.nan(mean_weight_per_day))] #remove the NaNs
    #days <- c(seq(from = 0, to = max(chickweight$Time), by = 2), 21)
    days  <- unique(chickweight$Time) #Same as above
    points(mean_weight_per_day~days, type = "l", col = chickcols[as.numeric(dietTypes)])
}
legend(1,300,paste("Diet",levels(chickweight$Diet)),bty='n',text.col=chickcols)
```
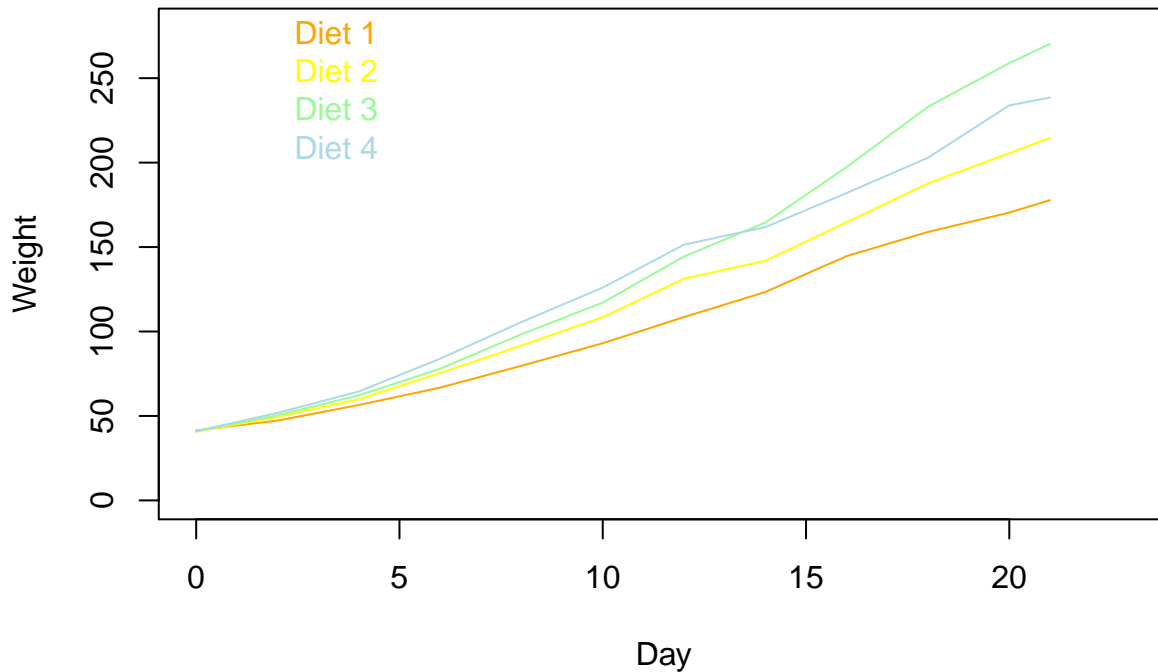


```
#Gaurav helped me come up with alternative and more elegant solutions.
#The first uses dplyr:
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
chickwts <- chickweight %>% group_by(Diet, Time) %>% summarise(weight = mean(weight))
plot(1, type = "n", xlim = c(0, 25), ylim = c(0, 280),
     xlab = "day", ylab = "weight", main = "Weight by day")

for(dietTypes in unique(chickwts$Diet)) {
```
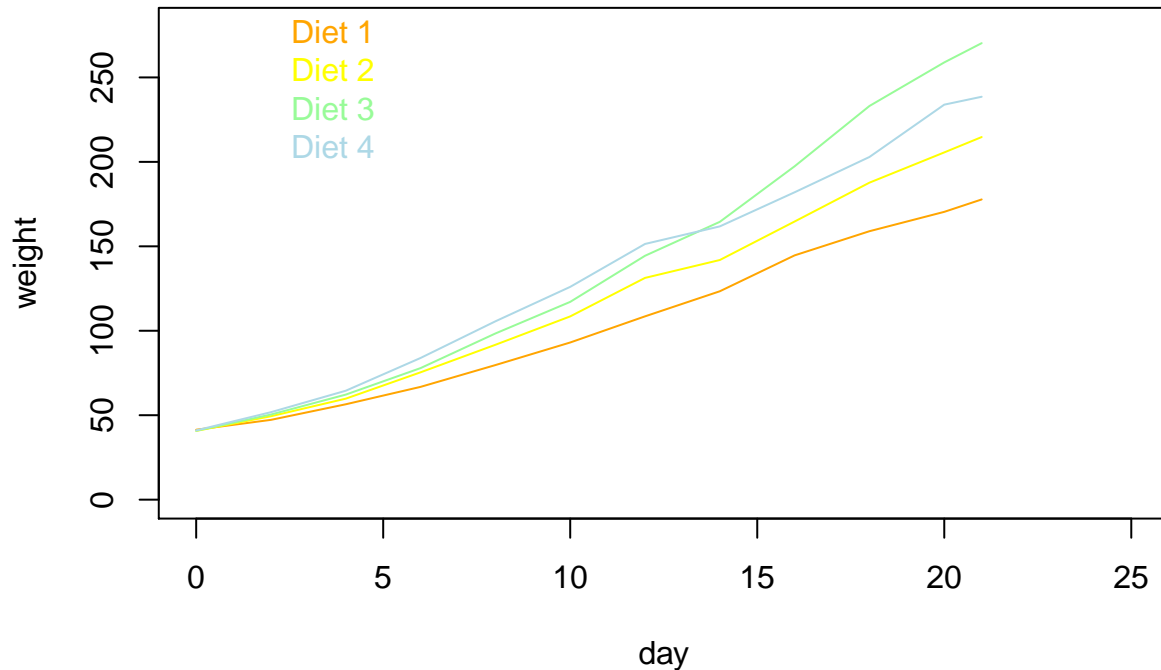
```
  current_diet <- chickwts %>% filter(Diet == dietTypes)
  lines(current_diet$weight~current_diet$Time,col=chickcols[as.numeric(dietTypes)])
}
legend(1,300,paste("Diet",levels(chickweight$Diet)),bty='n',text.col=chickcols)
```
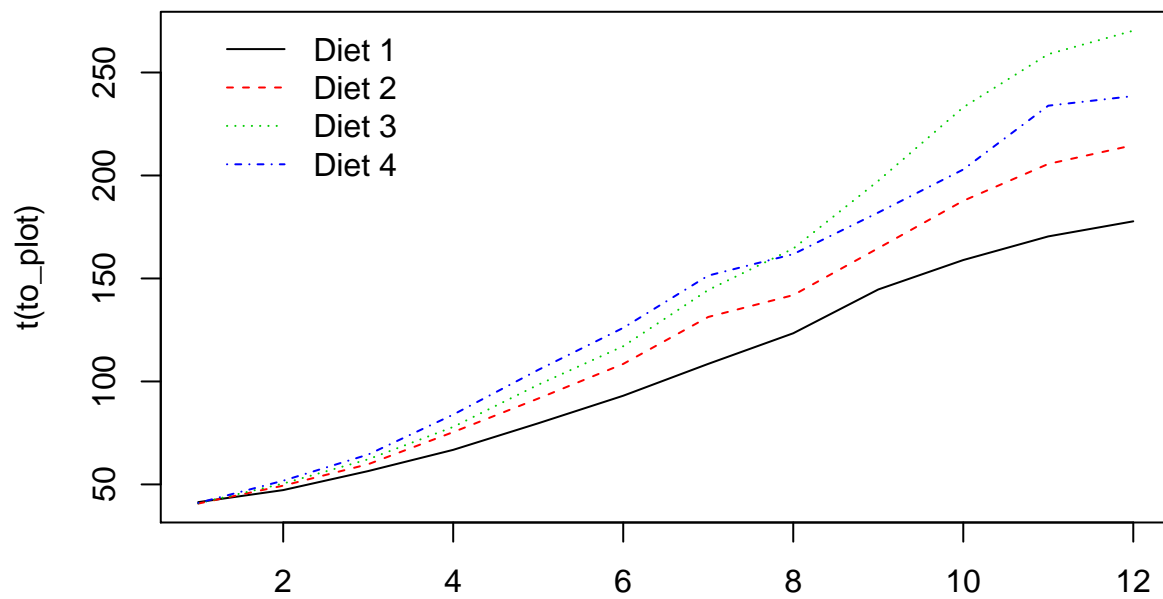
Weight by day


```
#The second uses aggregate:
xx <- aggregate(chickwts$weight, by = list(chickwts$Diet, chickwts$Time), FUN = mean)
to_plot <- matrix(xx$x, nrow = 4)
matplot(t(to_plot), type = "l")
legend(1,280,paste("Diet",levels(chickweight$Diet)),bty='n',lty=1:4,col=1:4)
#I knew the lty and col of the lines because the default in matplot is lty = 1:5 and col = 1:6


#The third uses ggplot2:
library(ggplot2)
```
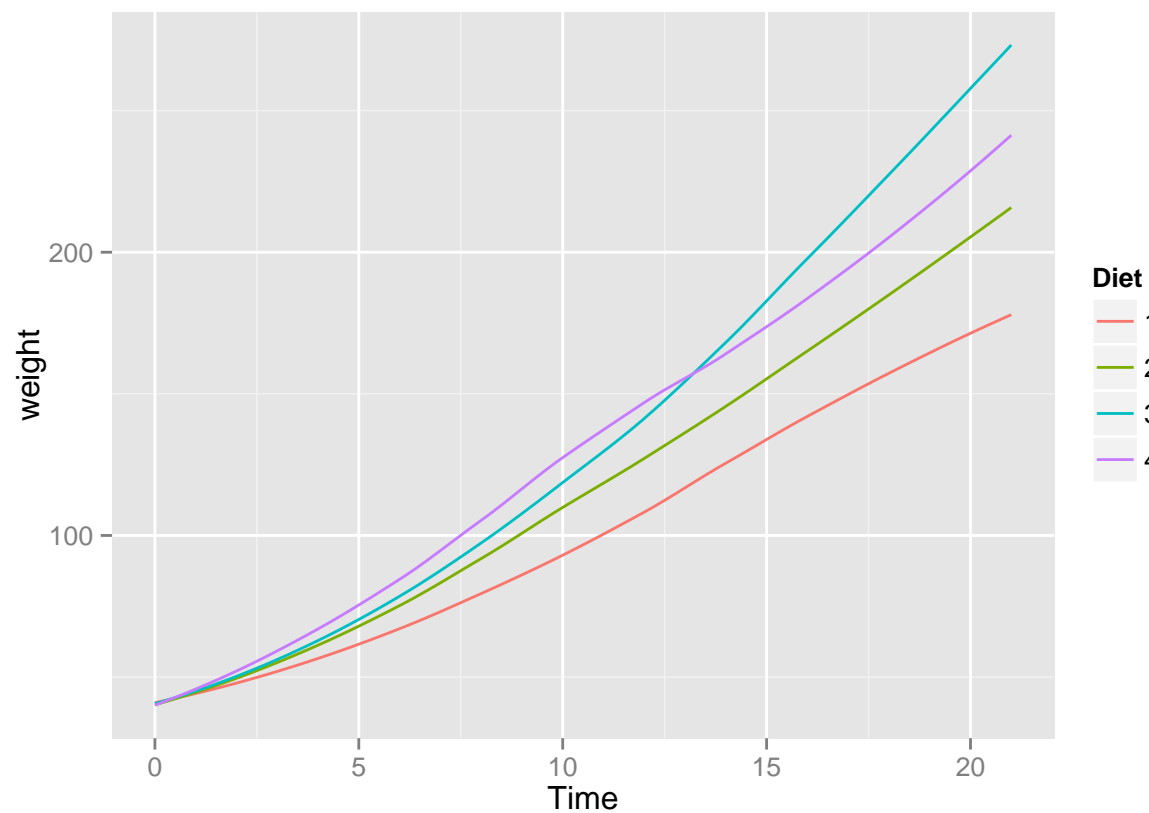
```
ggplot(chickwts, aes(Time, weight, color = Diet)) + geom_smooth(se = F)
```

```
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
```


5. Using a 'for' loop, plot the growth of each chick in a different color, all on the same plot.

```
par(mfrow=c(1,1))
chickweight$Time <- as.numeric(chickweight$Time)
chicki <- chickweight[chickweight$Chick==1,]
plot(chicki$Time,chicki$weight,type='l',ylim=c(40,380),xlab="Time (days)",ylab="Weight")
chicki <- NA
chickcols <- 1
for (i in 2:length(levels(chickweight$Chick))) {
  chicki <- chickweight[chickweight$Chick==i,]
  lines(chicki$Time,chicki$weight,col=chickcols)
  chickcols <- chickcols + 1
}
```
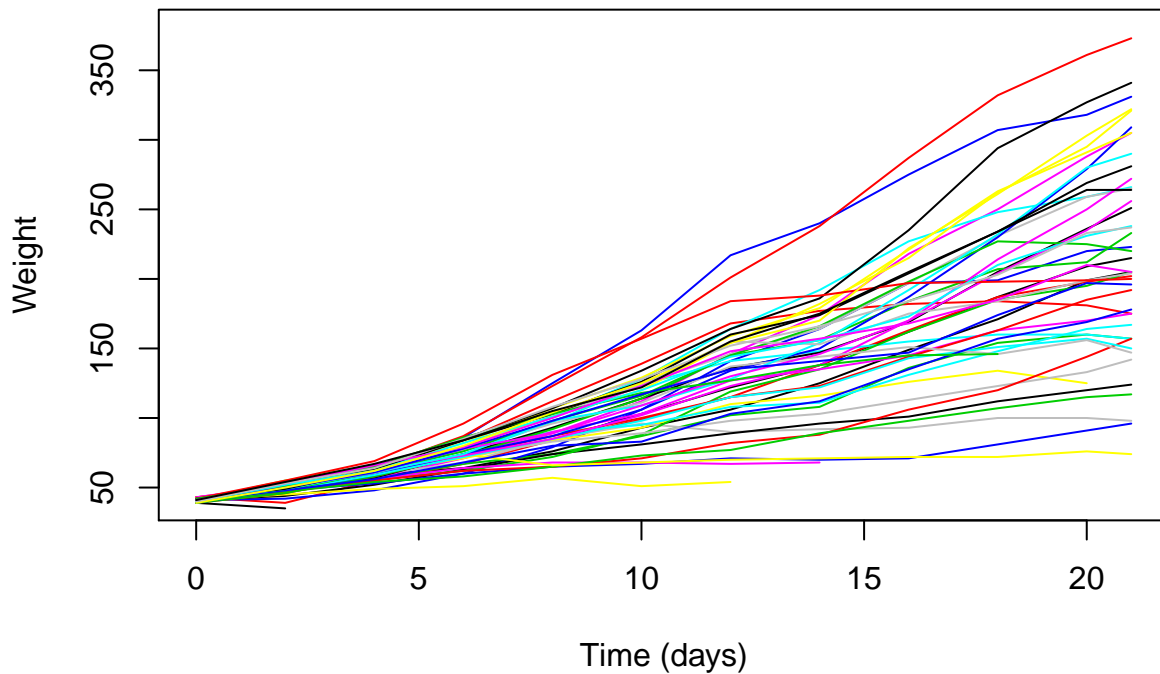

II. If we made it to the population growth part, using the model we wrote in class plot the population dynamics of 5 different populations, each with a different growth rate, all on the same plot.