Worksheet 1

2024-04-08

R Markdown

In this worksheet we will look at some brain data from some bats found in the Americas. The binomial names of each species are listed, alongside the family classifications. Some diet data are also included, indicating to which dietary guild each bat belongs: insects eater, fruit eater. Alongside brain size, which is depicted as a volume, there is also two other brain regions that were of evolutionary interest to the researchers: neocortex volume and olfactory bulb volume. The neocortex is a portion of the mammalian cerebral cortex involved in higher-order brain functions such as sensory perception, cognition, generation of motor commands, and spatial reasoning. The olfactory bulb is a region of the brain devoted to processing signals from the nose to identify certain scents in the air.

This worksheet asks you to produce some plots, identify trends in the data, and list some measures of central tendency.

This is the answer key. This key will demonstrate how I approached some of the questions, and show the exact scripts I used to build my plots or tables.

The following questions can be completed using either R or Excel.

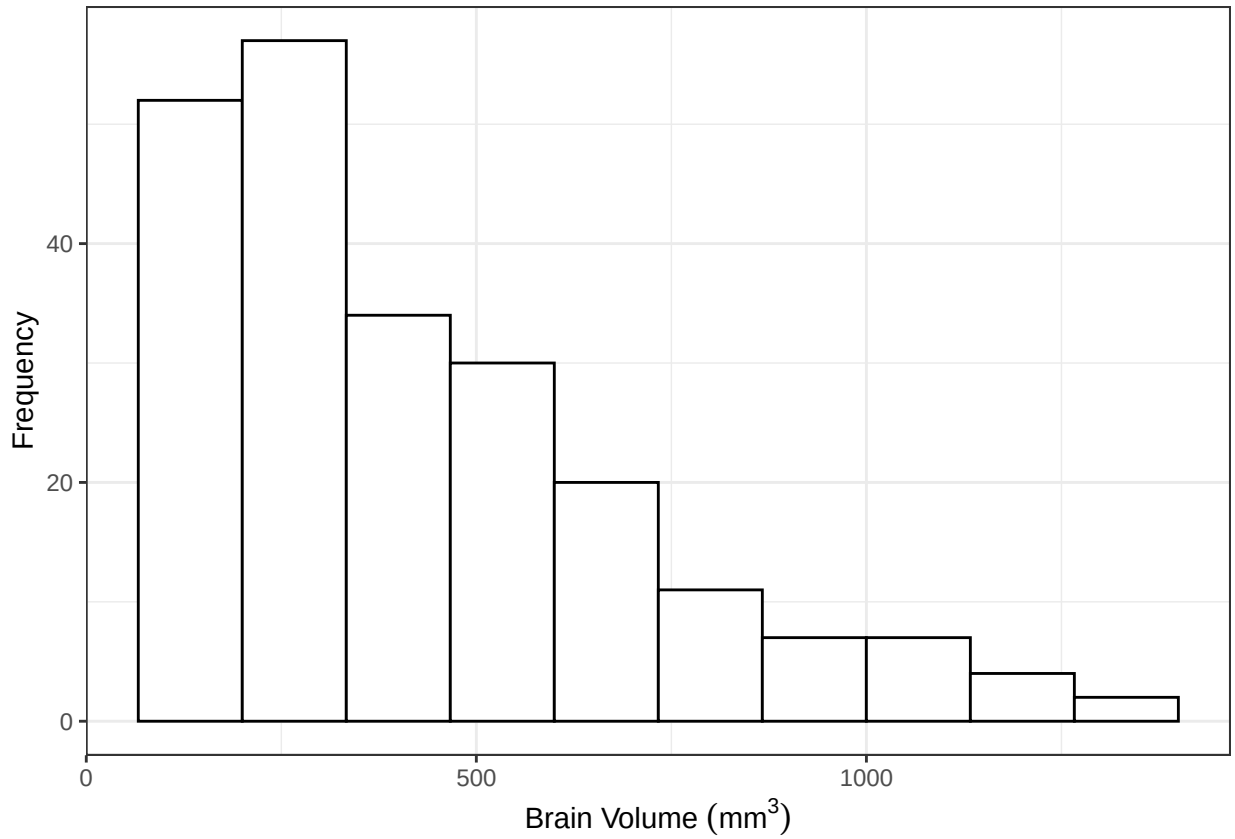
TASK 1

Open or import the brain volume data. Produce a histogram of the brain_size variable for the full data set, then produce a further two histograms with separate histograms for the fruit and insect feeding bats independently.

Note that I have already set my working directory as I am using R-Studio so I can set this using the environment rather than entering a line of code. I used the 'Session>Set Working Directory>Choose Directory...' pathway to set my directory.

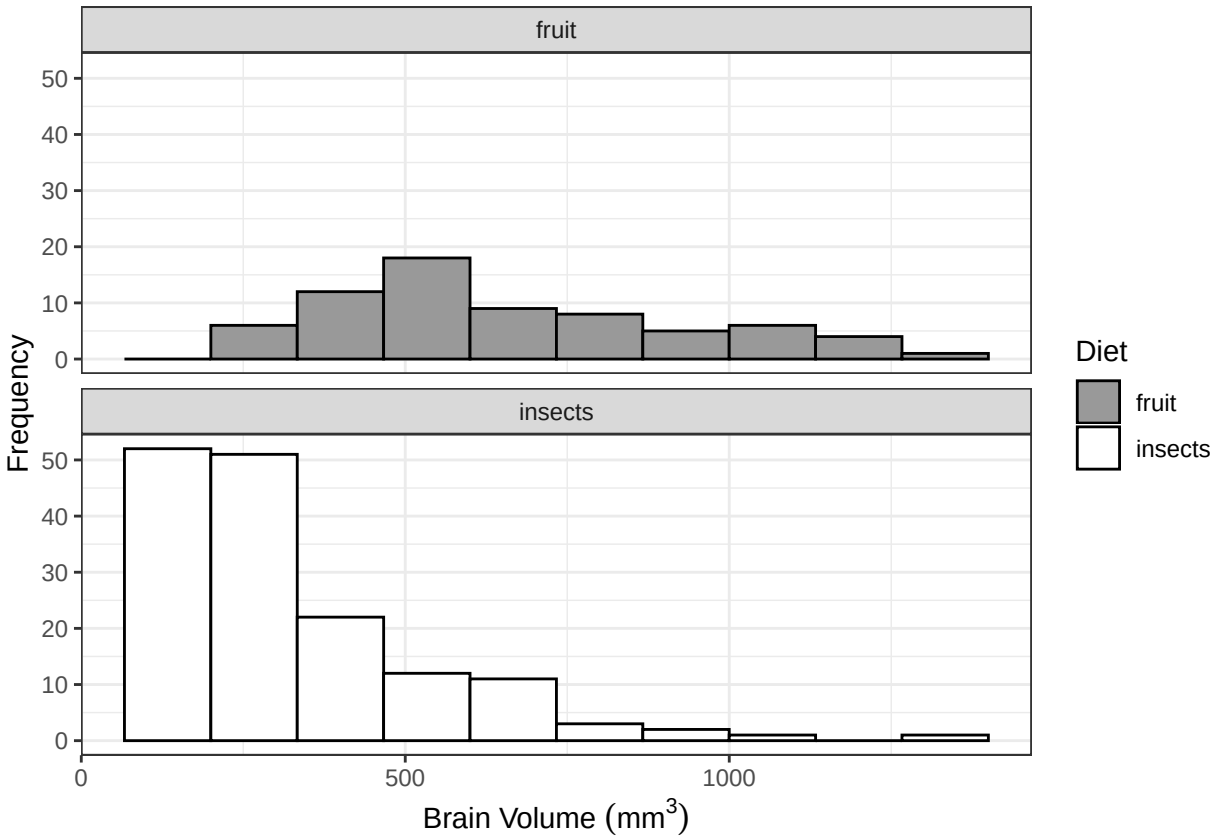
```
BatBrain<-read.csv("batbrains1.csv")

ggplot(BatBrain, aes(x=brain_size)) +
  geom_histogram(position="identity", bins = 10, color="black", fill="white") +
    theme_bw() + xlab(bquote('Brain Volume '(mm^3))) + ylab("Frequency")
```



You will find my combined data histogram above. Below you will find a pair of histograms where I have separated the data out based on their diets.

```
ggplot(BatBrain, aes(x=brain_size, color=Diet)) +
  geom_histogram(aes(fill=Diet), alpha=1, position="identity", bins = 10) +
  scale_fill_manual(values=c("#999999", "white")) +
  scale_color_manual(values=c("black", "black")) +
  theme_bw() + xlab(bquote('Brain Volume '(mm^3))) + ylab("Frequency") +
  facet_wrap(~Diet, nrow = 2)
```



Task 2

Examine all the plots you just produced. Do you see evidence of skew produced in your combined diet histograms and when the diets have become separated? Do you think one dietary category is driving the trends over the other category, or do they appear similar?

Task 3

Now calculate a group mean and median specific to each dietary category for the following variables brain_size, neocortex, olfactory_bulb. For example, the group mean for neocortex volume in the insect eating bats was 7.1mm3, while for fruit eating bats it was 30.5mm3. What about the other variables?

```
##Mean
aggregate(x = cbind(brain_size,neocortex,olfactory_bulb)~Diet,
          FUN="mean", data = BatBrain)
```

```
##      Diet brain_size neocortex olfactory_bulb
## 1   fruit   660.5261 188.37101      30.514493
## 2 insects   320.4335  72.06581       7.132903
```

```
##Median
aggregate(x = cbind(brain_size,neocortex,olfactory_bulb)~Diet,
          FUN="median", data = BatBrain)
```

```
##      Diet brain_size neocortex olfactory_bulb
## 1   fruit      584.0     161.4           27.5
## 2 insects      247.2      49.5            5.0
```

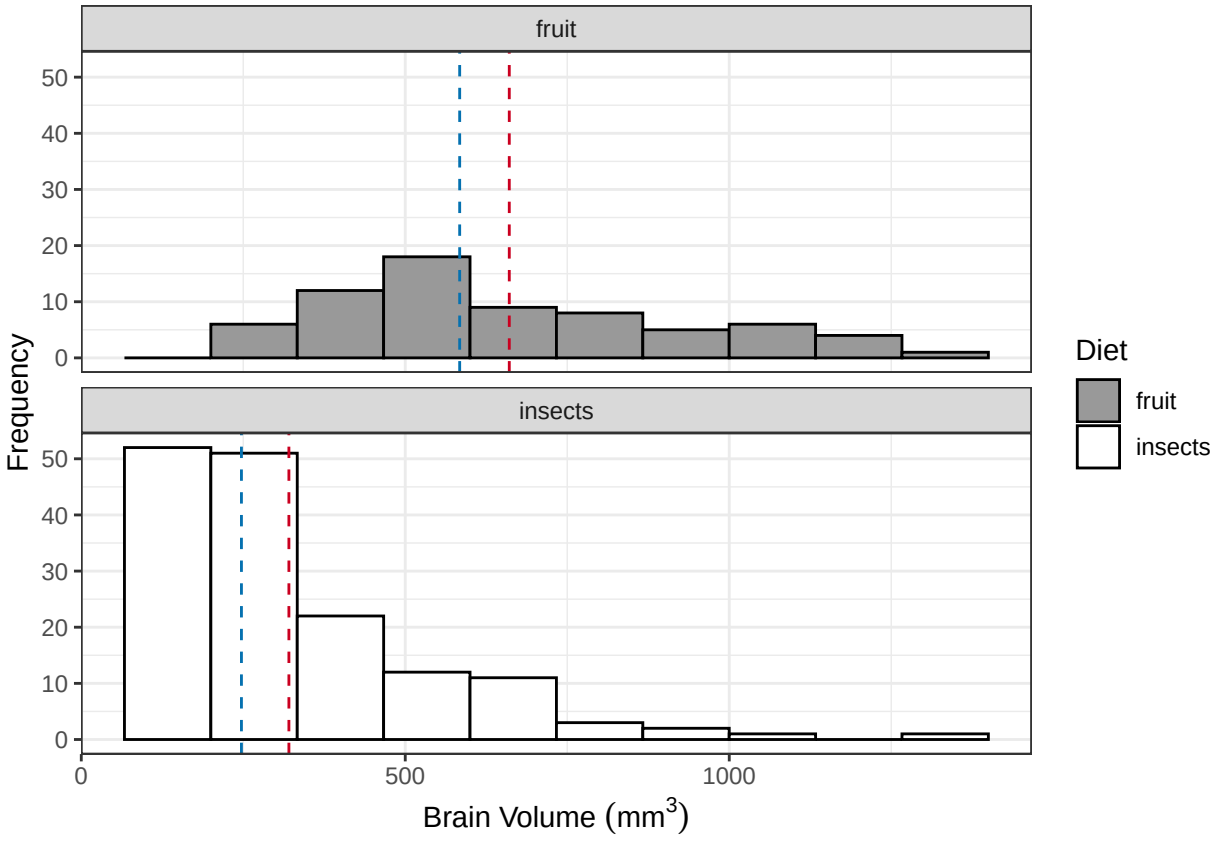
Task 4

What do you notice about the mean and median values for all your continuous traits? Are they similar? And, if not, are the means typically greater or smaller than the medians? What does this tell us about the distribution of our data?

I add an example of how the mean and median are different, plotting the mean as the red line and the median as the blue line on the histograms.

```
mu <- ddply(BatBrain, "Diet", summarise, grp.mean=mean(brain_size))
me <- ddply(BatBrain, "Diet", summarise, grp.mean=median(brain_size))

ggplot(BatBrain, aes(x=brain_size, color=Diet)) +
  geom_histogram(aes(fill=Diet), alpha=1, position="identity", bins = 10) +
  geom_vline(data=mu, aes(xintercept=grp.mean), color="#ca0020", linetype="dashed") +
  geom_vline(data=me, aes(xintercept=grp.mean), color="#0571b0", linetype="dashed") +
  scale_fill_manual(values=c("#999999", "white")) +
  scale_color_manual(values=c("black", "black")) +
  theme_bw() + xlab(bquote('Brain Volume '(mm^3))) + ylab("Frequency") +
  facet_wrap(~Diet, nrow = 2)
```



Task 5

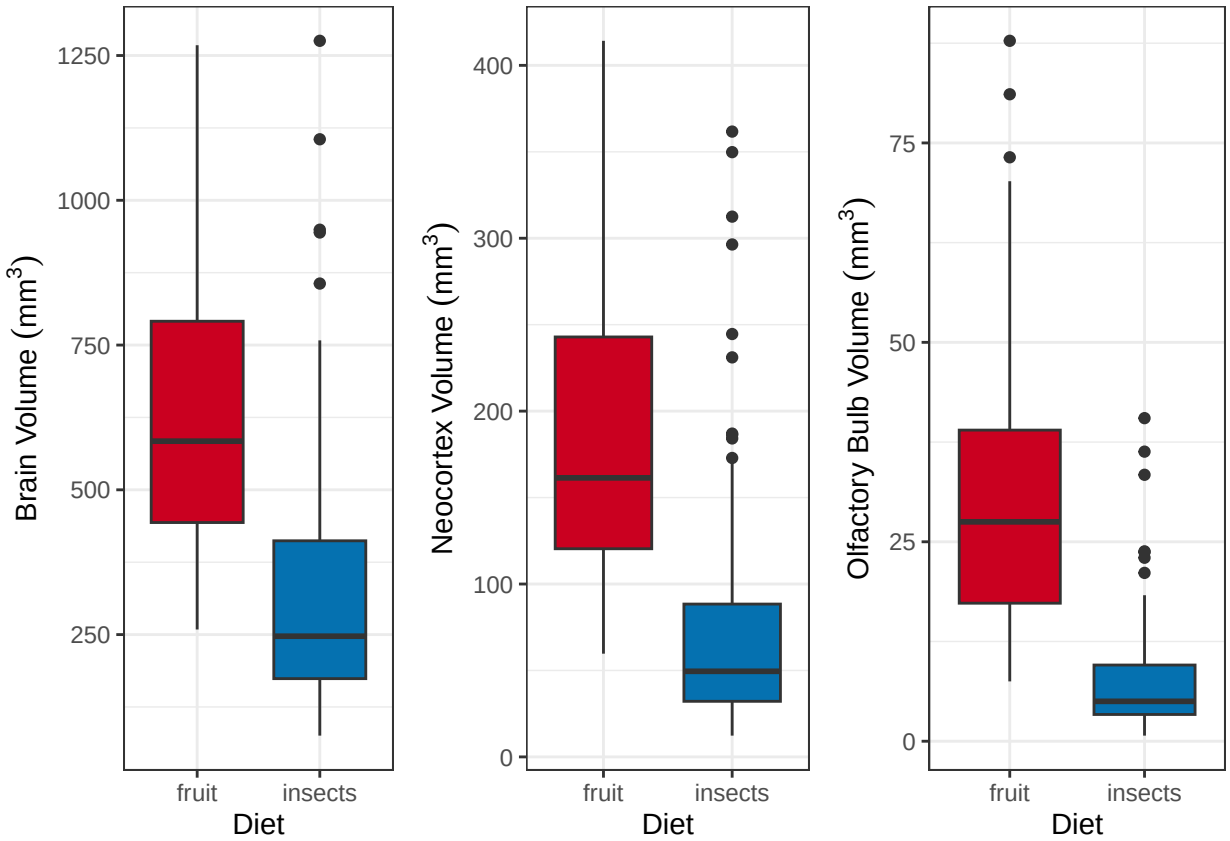
Produce a series of three boxplots for each of your continuous variables (brain volume, neocortex volume, olfactory bulb volume) with each of the three boxplots becoming split by diet (fruit, insects). What are some of the benefits of using a boxplot, that depicts the interquartile range, over the total range? What trends do you see in your boxplots - are the boxes and ranges similar for each variable between dietary categories?

```
A<-ggplot(BatBrain, aes(x=Diet, y=brain_size, fill=Diet)) +
  geom_boxplot() +
  scale_fill_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() + ylab(bquote('Brain Volume '(mm^3))) +
  guides(fill = "none")

B<-ggplot(BatBrain, aes(x=Diet, y=neocortex, fill=Diet)) +
  geom_boxplot() +
  scale_fill_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() + ylab(bquote('Neocortex Volume '(mm^3))) +
  guides(fill = "none")

C<-ggplot(BatBrain, aes(x=Diet, y=olfactory_bulb, fill=Diet)) +
  geom_boxplot() +
  scale_fill_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() + ylab(bquote('Olfactory Bulb Volume '(mm^3))) +
  guides(fill = "none")

plot_grid(A, B, C, nrow=1)
```

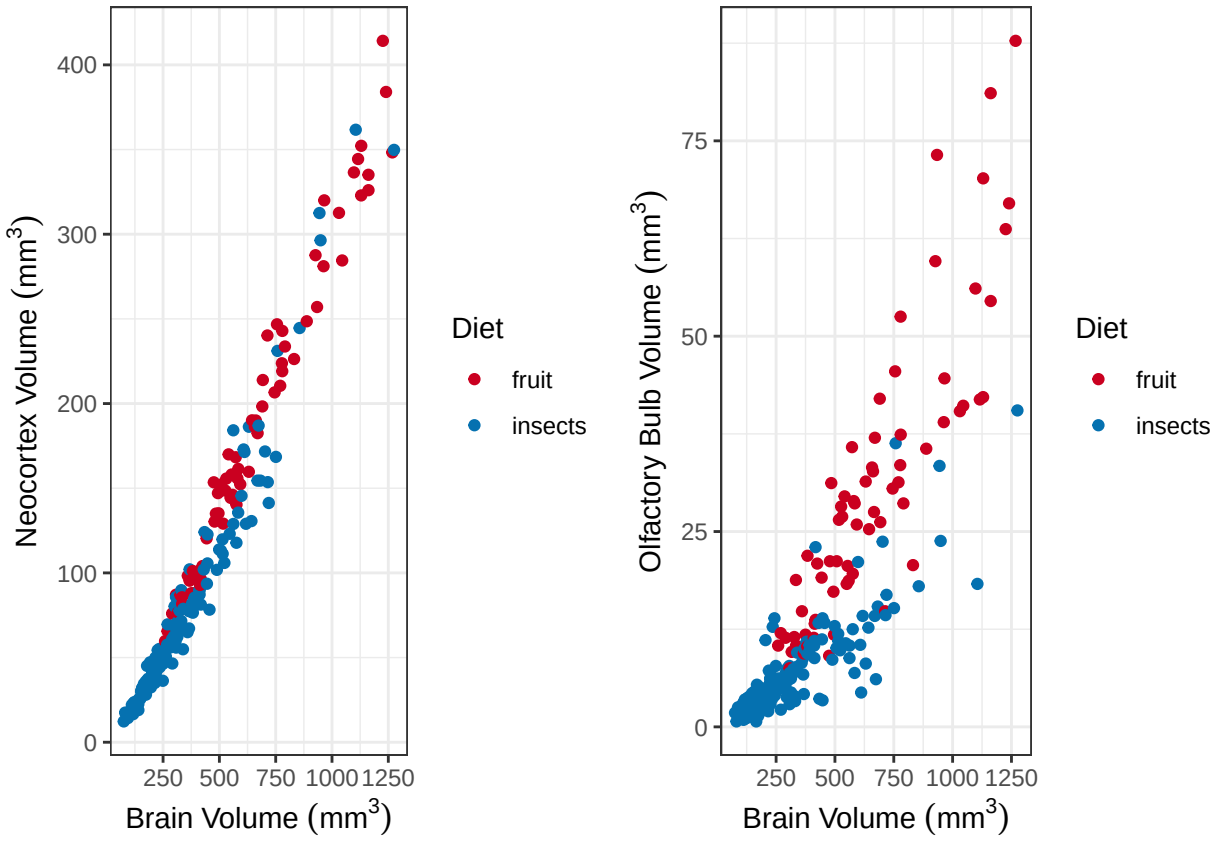


Task 6 Produce two scatter plots. Plot brain volume on the x-axis and neocortex volume on the y-axis, then for your second plot place brain volume on the x-axis and olfactory bulb volume on the y-axis. In each case, brain volume is classed as the what - dependent or independent variable? Describe the trend in the variation of your data as brain size increases.

```
K<-ggplot(BatBrain, aes(x=brain_size, y=neocortex, color=Diet)) +
  geom_point() +
  scale_color_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() +
  xlab(bquote('Brain Volume '(mm^3))) + ylab(bquote('Neocortex Volume '(mm^3)))

L<-ggplot(BatBrain, aes(x=brain_size, y=olfactory_bulb, color=Diet)) +
  geom_point() +
  scale_color_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() +
  xlab(bquote('Brain Volume '(mm^3))) + ylab(bquote('Olfactory Bulb Volume '(mm^3)))

plot_grid(K, L, nrow=1)
```



Task 7 Calculate standard deviation for the olfactory bulb for bats in each dietary category. Which dietary category exhibits the greatest standard deviation?

```
##Mean
aggregate(x = cbind(brain_size,neocortex,olfactory_bulb)~Diet,
          FUN="sd", data = BatBrain)
```

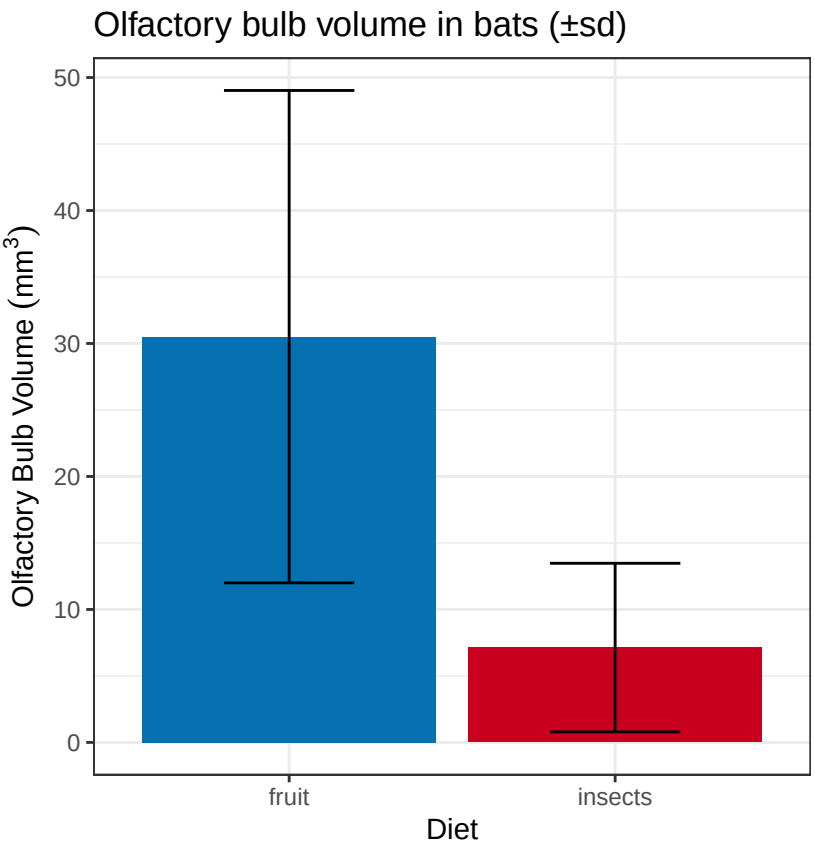
```
##      Diet brain_size neocortex olfactory_bulb
## 1   fruit   276.0293  89.44812      18.513066
## 2 insects   213.1053  63.05497       6.339129
```

Task 8 #Removed, this was too difficult at this point## Now produce a bar plot to reflect the differences in standard deviation. Graph the olfactory bulb volume for both the fruit and insect dietary categories, and add the standard deviation as an error bar to each bar.

```
my_sd_Brain <- BatBrain %>%
  group_by(Diet) %>%
  summarise(
    n=n(),
    mean=mean(olfactory_bulb, na.rm = T),
    sd=sd(olfactory_bulb, na.rm = T)
  )

# Standard error
Brain_Olfactory <- ggplot(my_sd_Brain, aes(x=Diet, y=mean, fill=factor(Diet))) +
  geom_bar(stat="identity", position='dodge', alpha=1) +
  scale_fill_manual(values=c("#0571b0","#ca0020"), name="Diet",
                    labels=c("Fruit", "Insects")) +
  geom_errorbar(aes(x=Diet, ymin=mean-sd, ymax=mean+sd),
                position = position_dodge(.9), width=0.4, colour="black", alpha=1) +
  ggtitle("Olfactory bulb volume in bats (±sd)") +
  labs(y = bquote('Olfactory Bulb Volume '(mm^3)), x = "Diet", fill = "Diet") +
  theme_bw() + theme(aspect.ratio=1, legend.position = "none") + coord_fixed()

Brain_Olfactory
```



Task 9 Convert your neocortex volume and olfactory bulb volume data to a Z-score and plot them as two

boxplots split by diet, similar to our output from Task 5. Transforming them to Z-scores allows us to visualize and compare the spread of the data more easily, despite their initial difference in scale. Which variable, neocortex volume or olfactory bulb volume, exhibits greater overall variation?

```
z_score1 <- (BatBrain$neocortex-mean(BatBrain$neocortex))/sd(BatBrain$neocortex)
BatBrain$neocortexZ <- z_score1

z_score2 <- (BatBrain$olfactory_bulb-mean(BatBrain$olfactory_bulb))/sd(BatBrain$olfactory_bulb)
BatBrain$olfactory_bulbZ <- z_score2

R<-ggplot(BatBrain, aes(x=Diet, y=neocortexZ, fill=Diet)) +
  geom_boxplot() +
  scale_fill_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() + ylim(-2, 6) + ylab('Neocortex Volume Z-Score')

Q<-ggplot(BatBrain, aes(x=Diet, y=olfactory_bulbZ, fill=Diet)) +
  geom_boxplot() +
  scale_fill_manual(values=c("#ca0020", "#0571b0")) +
  theme_bw() + ylim(-2, 6) + ylab('Olfactory Bulb Volume Z-Score')

plot_grid(R, Q, nrow=1)
```