Not Project 1

Caleb King

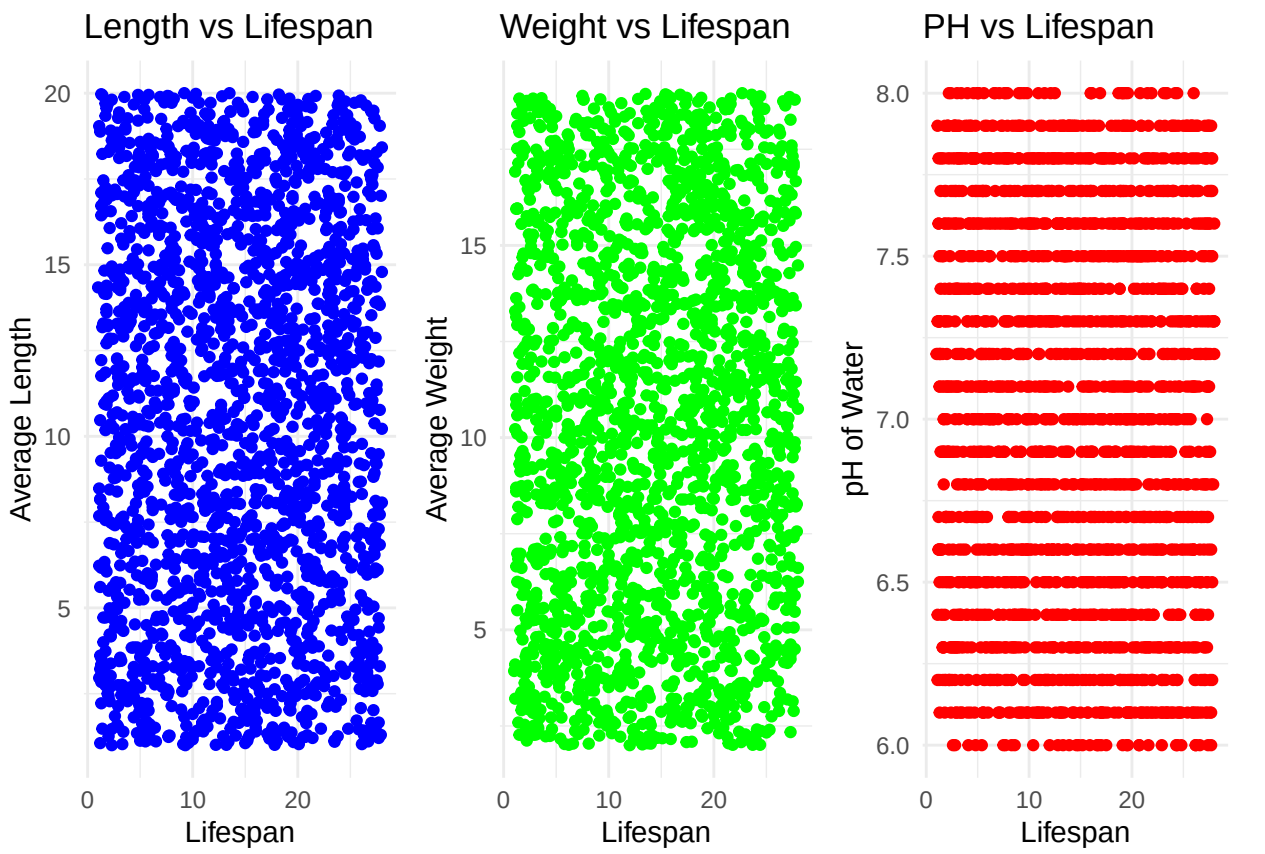
Summary of cleaned data

```
library(TestingPackage)
library (ggplot2)
summary(fishData)
```

```
##        id           avg_length      avg_weight                habitat
##  Min.   :   1.0   Min.   : 1.00   Min.   : 2.000   idlewater       :380
##  1st Qu.: 494.8   1st Qu.: 5.85   1st Qu.: 6.128   lakes           :417
##  Median : 988.5   Median :10.68   Median :10.455   ponds           :396
##  Mean   : 995.1   Mean   :10.56   Mean   :10.449   rivers          :391
##  3rd Qu.:1499.2   3rd Qu.:15.17   3rd Qu.:14.700   slowmovingwaters:392
##  Max.   :2000.0   Max.   :20.00   Max.   :18.960
##
##   ph_of_water                    color        Gender       life_span
##  Min.   :6.000   Green_and_Chocolate:254   Female:1007   Min.   : 1.000
##  1st Qu.:6.500   Red_and_Silver     :239   Male  : 969   1st Qu.: 7.875
##  Median :7.000   Yellow             :239                 Median :14.400
##  Mean   :7.015   Orange             :236                 Mean   :14.389
##  3rd Qu.:7.500   Black_and_Orange   :212                 3rd Qu.:20.900
##  Max.   :8.000   Reddish_Orange     :211                 Max.   :28.000
##                  (Other)            :585
```

lifespan against each numerical variable (scatterplot)

```
library(gridExtra)
plot1 <- ggplot(data = fishData, aes(x = life_span, y = avg_length)) +
  geom_point(color = "blue") +
  labs(x = "Lifespan", y = "Average Length") +
  ggtitle("Length vs Lifespan") +
  theme_minimal()

plot2 <- ggplot(data = fishData, aes(x = life_span, y = avg_weight)) +
  geom_point(color = "green") +
  labs(x = "Lifespan", y = "Average Weight") +
  ggtitle("Weight vs Lifespan") +
  theme_minimal()

plot3 <- ggplot(data = fishData, aes(x = life_span, y = ph_of_water)) +
  geom_point(color = "red") +
  labs(x = "Lifespan", y = "pH of Water") +
  ggtitle("PH vs Lifespan") +
  theme_minimal()

# Combine plots into one output
combined_plots <- grid.arrange(plot1, plot2, plot3, ncol = 3)
```




```
# Output the combined plot
print(combined_plots)
```

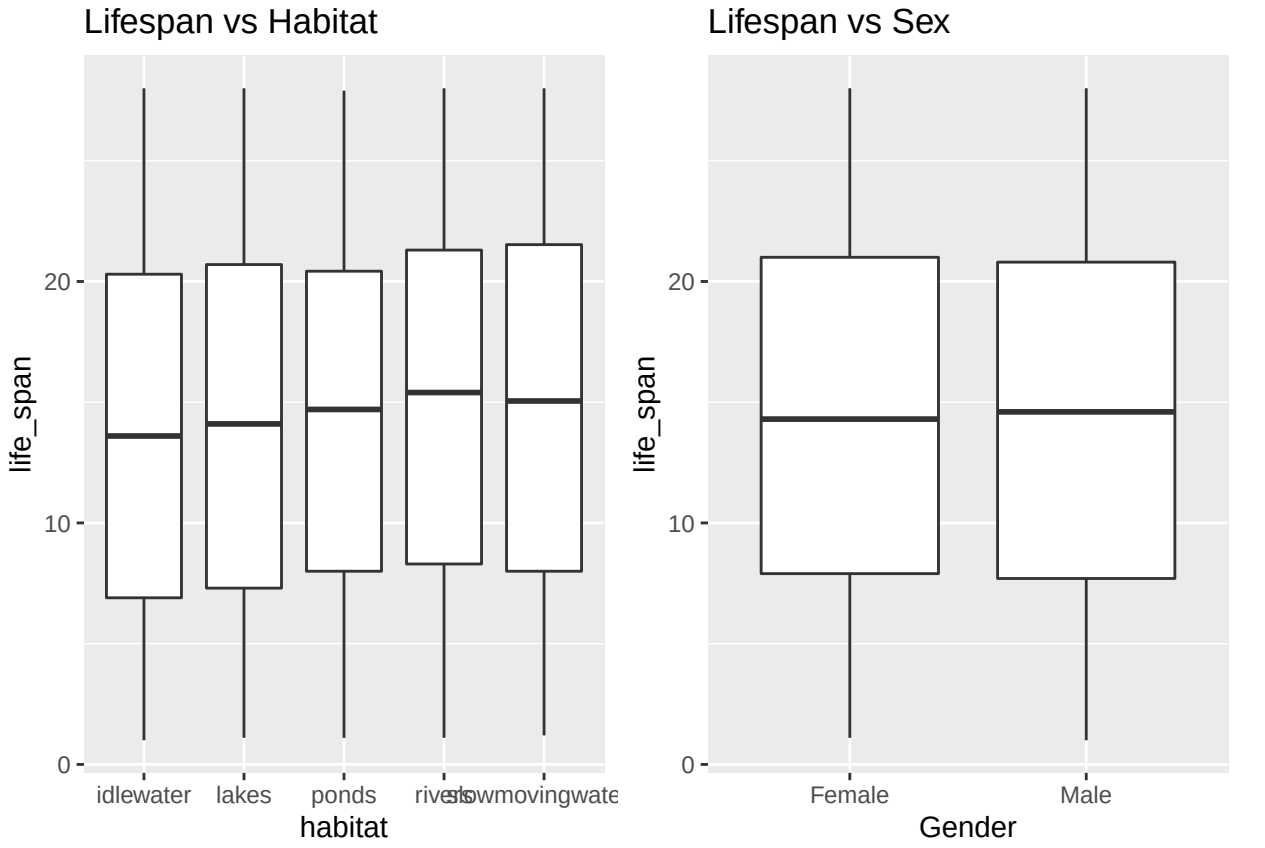
```
## TableGrob (1 x 3) "arrange": 3 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
```

```
#ggplot(cleanData, aes(x = habitat, fill = color)) +
#  geom_bar() +
#  labs(title = "Distribution of Fish Habitat",
#       x = "Habitat",
#       y = "Number of Fish")
```

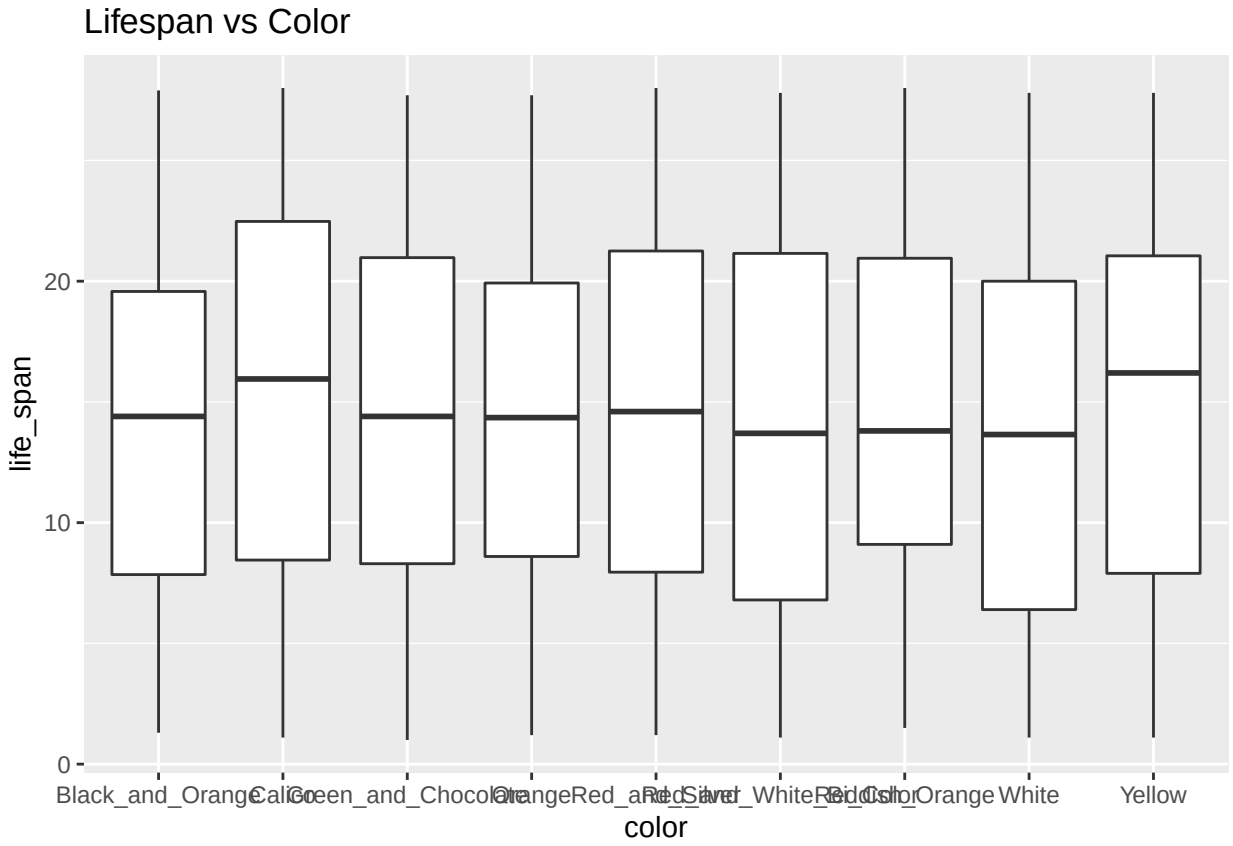
Lifespan against categorical variables (boxplot)

```
plot4 <- ggplot(fishData, aes(x = habitat, y = life_span)) +
  geom_boxplot() +
  labs(title = "Lifespan vs Habitat")

plot5 <- ggplot(fishData, aes(x = color, y = life_span)) +
  geom_boxplot() +
  labs(title = "Lifespan vs Color")

plot6 <- ggplot(fishData, aes(x = Gender, y = life_span)) +
  geom_boxplot() +
```

```
  labs(title = "Lifespan vs Sex")

combined_plots <- grid.arrange(plot4, plot6, ncol = 2)
```




```
print(plot5)
```




```
print(combined_plots)
```

```
## TableGrob (1 x 2) "arrange": 2 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
```

Lifespan of each habitat

```
#boxplot(life_span ~ habitat, data = cleanData, main = "Lifespan for Each Habitat")
```

PH of water for each habitat

```
#ggplot(cleanData, aes(x = ph_of_water, y = ..density..)) +
#  geom_histogram(fill = "cornsilk", color = "grey60") +
#  facet_grid(habitat ~ .) +
#  geom_density() +
#  labs(title = "Histogram of PH of Water for Each Habitat",
#       x = "PH of Water",
#       y = "Density of Fish")
```