Module 3-1

Keaton Wilson

2/6/2020

Introduction to the Problem and Summarizing Data

The challenge

Part of the food that folks are being fed with are fish that are caught, not farmed (as in the previous module). These fish are caught by members of team antarctica, but there is extreme risk involved (think Deadliest Catch, but colder).

One of the main hazards to these teams are leopard seals - they're apex predators in this ecosystem, and can cause serious injury, especially when large schools of fish are involved (a researcher was killed by a Leopard Seal in Antarctica in 2003).

One way we've worked to remove some of this danger is by working with large mammal researchers to place radio collars on a number of seals that live in areas that are typically fished. Fishing boats are equipped with radios that can detect the presence of seals in the area and avoid high-risk sites.

In recent months, a problem has arisen. Some of the collars are failing, leading to some very close calls - one team member was pulled into the water after attempting to untangle a net when a seal lunged at some fish trapped in the net.

The large mammal team (they work on orcas as well as leopard seals) replaces collars frequently, based on the manufacters reccomendations for battery life and general wear and tear, but it seems as if some of the collars are dying earlier than expected and putting our team in danger.

We've been tasked with determining why the units are failing and if we can tie it to a particular manufacturer (we've sourced radio collars from two different companies).

The Data

Our data on collars can be accessed here: https://tinyurl.com/sp7b25x

Let's explore and talk about the data a bit:
1. Two manufacturers
2. Battery life - The average number of days a particular collar lasts (this is recorded in the unit and stored when the battery dies)
3. Signal distance - the maximum signal distance that a particular collar was recorded at.
4. Fail - collars that have failed in the past (e.g. they've been recovered from seals that were noticed by the team but that didn't ping the radio equipment).

Group Challenge 1 - Summarizing

We covered how to do grouped summaries in R in Module 1 (and practiced a bit in module 2), while also covering how to write your own custom functions in module 2. It's time to put this together. Write your

own custom function that does the following:
1. Summarizes the data by calculating the mean, min and max for battery life and signal distance, while counting up the number of failures. You'll want to do this for both manufacturers.

```
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ----------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
collars = read_csv("https://tinyurl.com/sp7b25x")
```

```
## Parsed with column specification:
## cols(
##   collar_id = col_double(),
##   maker = col_character(),
##   battery_life = col_double(),
##   signal_distance = col_double(),
##   fail = col_double()
## )
```

```
summary_1 = collars %>%
  mutate(fail = as.logical(fail),
         maker = as.factor(maker)) %>%
  select(-collar_id) %>%
  group_by(maker) %>%
  summarize_if(is.numeric, list(min = min, max = max, mean = mean)) %>%
  select(-contains("collar_id"))

summary_2 = collars %>%
  mutate(fail = as.logical(fail),
         maker = as.factor(maker)) %>%
  select(-collar_id) %>%
  group_by(maker) %>%
  summarize_if(is.logical, sum)

final = bind_cols(summary_1, summary_2) %>%
  select(-maker1)
```

Visualizing the Data - an introduction to a better way to visualize the data

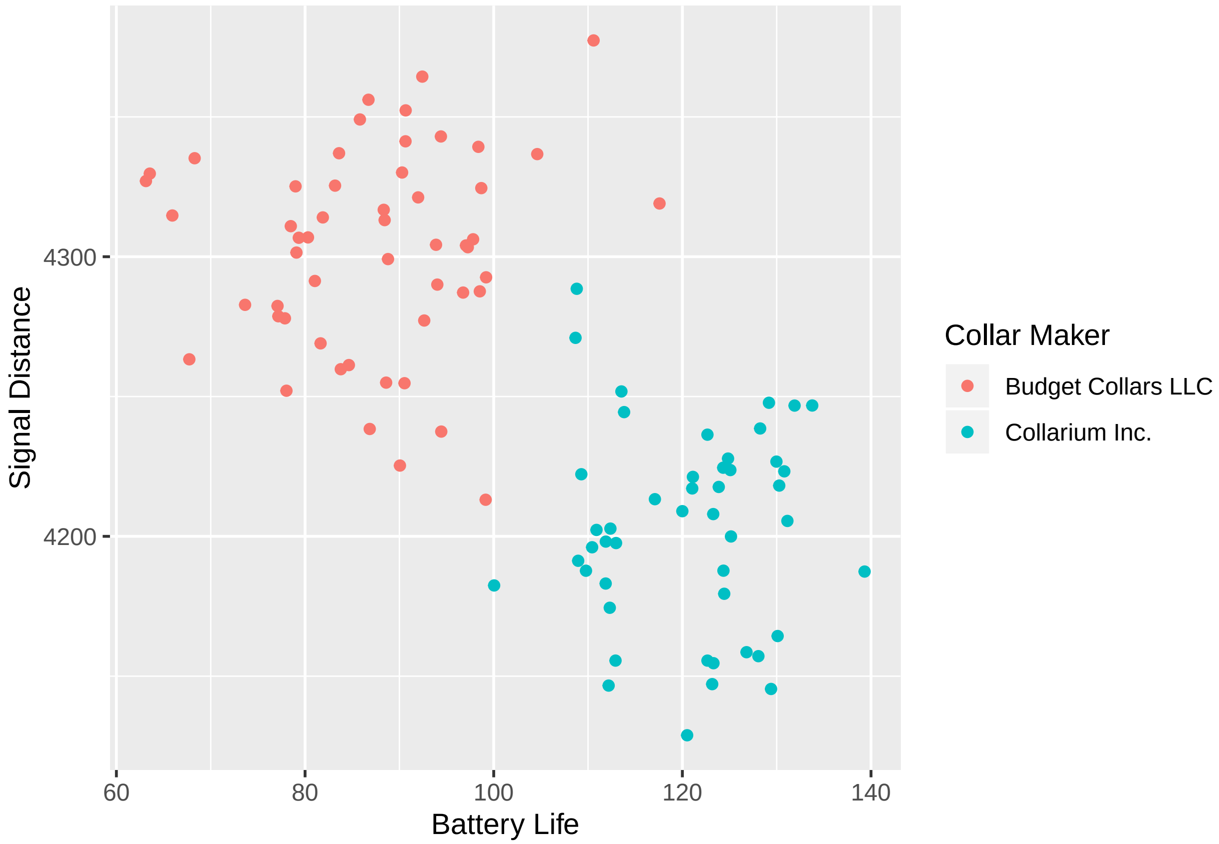
So far, we've used the base R plotting syntax, but there is another way. It's part of the tidyverse and its from a package called ggplot2.

Instead of explaining the syntax to you, we're going to do an in-class reading assignment.

Article to read

Spend 10 minutes reading the article above (take notes!). When the time is up, I'll signal you to get into your groups and discuss the article. There is a lot of material here - make sure that you cover what the grammar of graphics is, what the different components are, and how they relate to code structure and syntax. We'll be iteratively building some ggplot code on the data above, and I'll be asking groups to explain each part as we go - so be prepared!

```
ggplot(collars, aes(y = signal_distance, x = battery_life, col = maker)) +
  geom_point() +
  xlab("Battery Life") +
  ylab("Signal Distance") +
  scale_color_discrete(name = "Collar Maker")
```


Signal Distance
Battery Life
Collar Maker
● Budget Collars LLC
● Collarium Inc.