

Exploratory Data Analysis - Summary Report

1. Data Overview

Table 1: Composition of the sample

```
library(tidyverse)
library(palmerpenguins)
data(penguins)
summary(penguins)
```

species	island	bill_length_mm	bill_depth_mm
Adelie :152	Biscoe :168	Min. :32.10	Min. :13.10
Chinstrap: 68	Dream :124	1st Qu.:39.23	1st Qu.:15.60
Gentoo :124	Torgersen: 52	Median :44.45	Median :17.30
		Mean :43.92	Mean :17.15
		3rd Qu.:48.50	3rd Qu.:18.70
		Max. :59.60	Max. :21.50
		NA's :2	NA's :2

flipper_length_mm	body_mass_g	sex
Min. :172.0	Min. :2700	female:165
1st Qu.:190.0	1st Qu.:3550	male :168
Median :197.0	Median :4050	NA's : 11
Mean :200.9	Mean :4202	
3rd Qu.:213.0	3rd Qu.:4750	
Max. :231.0	Max. :6300	
NA's :2	NA's :2	

According to the Table 1 the majority of penguins are Adelie and the majority of penguins are caught from Biscoe island. Except, in the variables species and island there are 2 missing values in other variables.

2. Composition of the sample

```
ggplot(penguins) + geom_bar(aes(species))
```

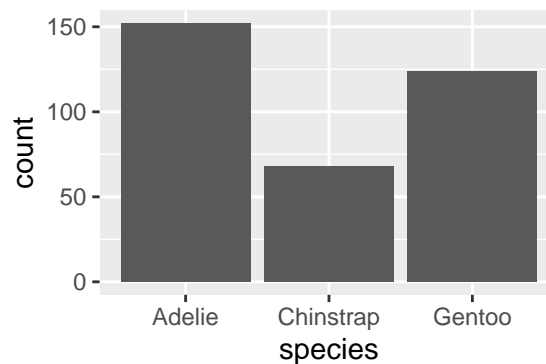


Figure 1: Composition of sample by Species

Sometimes, you will see row names as a variable. Then you will get multiple bars like this. Here is an example. See Figure 2. Do not include those types of charts to the report.

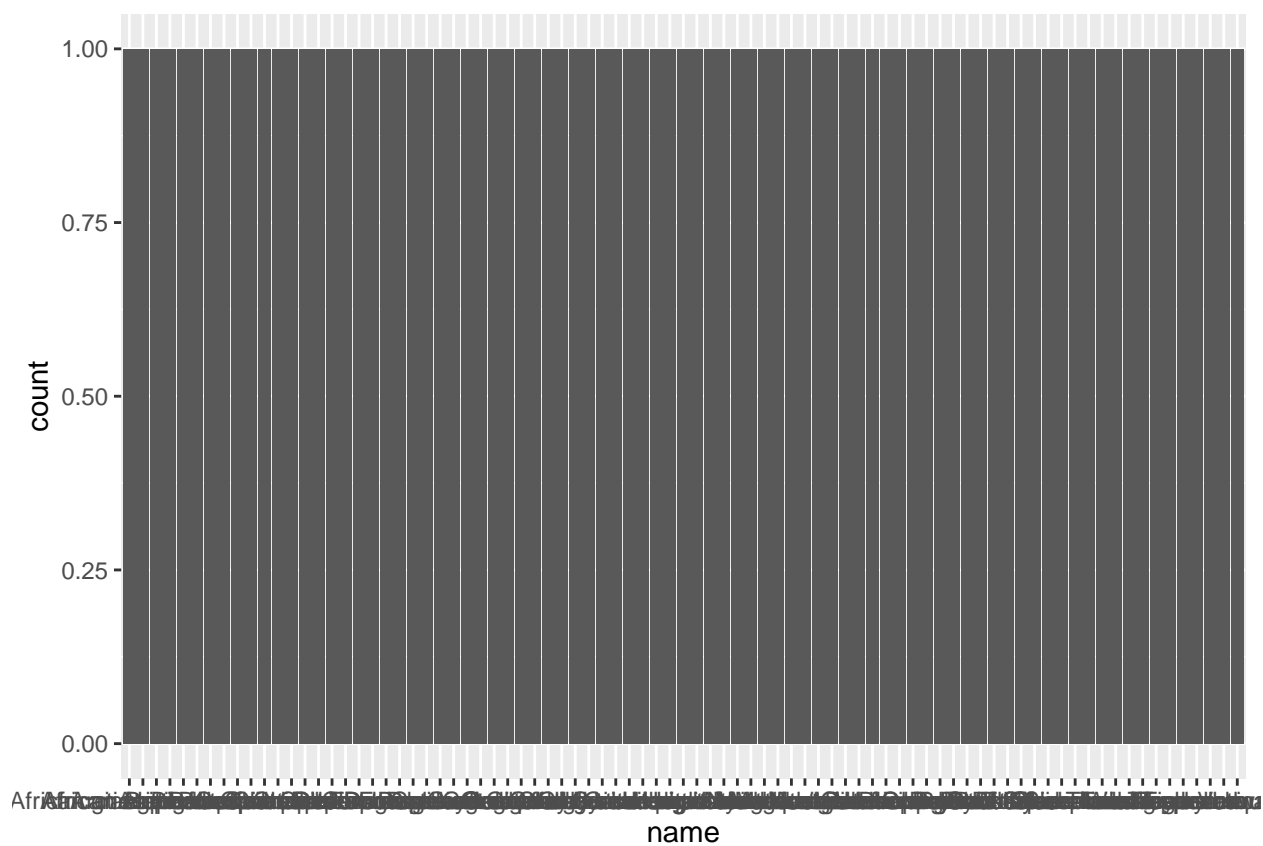


Figure 2: Do not include these types of graphs