

## R bootcamp exercise - Antogil Bay fisheries data

Daniel Viana and Jessica Zamborain Mason

October, 2022

We are going to work with a dataset that contains post-harvest data for 5 villages in Antongil Bay, Madagascar. This data was collected with fishers between 2016 and 2017 and provides information such as species, catch (kg), number of fishers and fishing duration (effort in hours). In addition, it contains information on how much of the catch was destined to human consumption and how much was sold. Fishing gears are represented by numbers (net = 1, line = 2, palangra = 3, spear = 4, vovo = 5, hand net = 6, other = 7, multiple = 8).

- 1) Read explore the data ("CPUE\_data\_Rcourse.csv"). How many variables does the dataset have? What information it contains?
- 2) Do a simple plot of the catch records through time.
- 3) What is the total catch per village? Create a figure to show this.
- 4) What is the cath per fishing gear in each village?
- 5) What are the top 3 species caught in the entire bay?
- 6) What is the average monthly CPUE (catch per unit effort) of the top 3 species caught in the bay? Is there any seasonality on the CPUE throughout the year? What does this mean?
- 7) for each village, what proportion of total catch is consumed vs sold?
- 8) What other interesting analysis can you do with the data? What other interesting questions can you ask? try to do it!

\*Answers

- 1) Read explore the data ("CPUE\_data\_Rcourse.csv"). How many variables does the dataset have? What information it contains?

```
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse
1.3.0 —
```

```
## ✓ ggplot2 3.3.2      ✓ purrr  0.3.3
## ✓ tibble  2.1.3      ✓ dplyr  0.8.4
## ✓ tidyr   1.0.2      ✓ stringr 1.4.0
## ✓ readr   1.3.1      ✓ forcats 0.5.0
```

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

```
dta <- read_csv("~/Rbootcamp_Madagascar/data/CPUE_data_Rcourse.csv")
```

```
## Parsed with column specification:
## cols(
##   fish_name = col_character(),
##   scientific_name = col_character(),
##   family = col_character(),
##   genus = col_character(),
##   gear = col_double(),
##   fisher_ID = col_double(),
##   date = col_character(),
##   month = col_double(),
##   year = col_double(),
##   village = col_double(),
##   time_decimal = col_double(),
##   total_weight = col_double(),
##   number_fish = col_double(),
##   total_sold = col_double(),
##   total_consumed = col_double()
## )
```

```
head(dta)
```

```
## # A tibble: 6 x 15
##   fish_name scientific_name family genus gear fisher_ID date month year
##   <chr>      <chr>          <chr> <chr> <dbl>      <dbl> <chr> <dbl> <dbl>
## 1 laoko      <NA>          <NA> <NA>  1         28 6/17...  6  2017
## 2 laoko      <NA>          <NA> <NA>  8         9 11/2... 11  2016
## 3 laoko      <NA>          <NA> <NA>  1         28 9/9/...  9  2016
## 4 laoko      <NA>          <NA> <NA>  8         9 5/15...  5  2017
## 5 laoko      <NA>          <NA> <NA>  1         13 11/2... 11  2016
## 6 laoko      <NA>          <NA> <NA>  7         24 7/19...  7  2016
## # ... with 6 more variables: village <dbl>, time_decimal <dbl>,
## #   total_weight <dbl>, number_fish <dbl>, total_sold <dbl>,
## #   total_consumed <dbl>
```

```
names(dta)
```

```
## [1] "fish_name"      "scientific_name" "family"          "genus"
## [5] "gear"           "fisher_ID"       "date"            "month"
## [9] "year"           "village"         "time_decimal"    "total_weight"
## [13] "number_fish"    "total_sold"      "total_consumed"
```

```
str(dta)
```

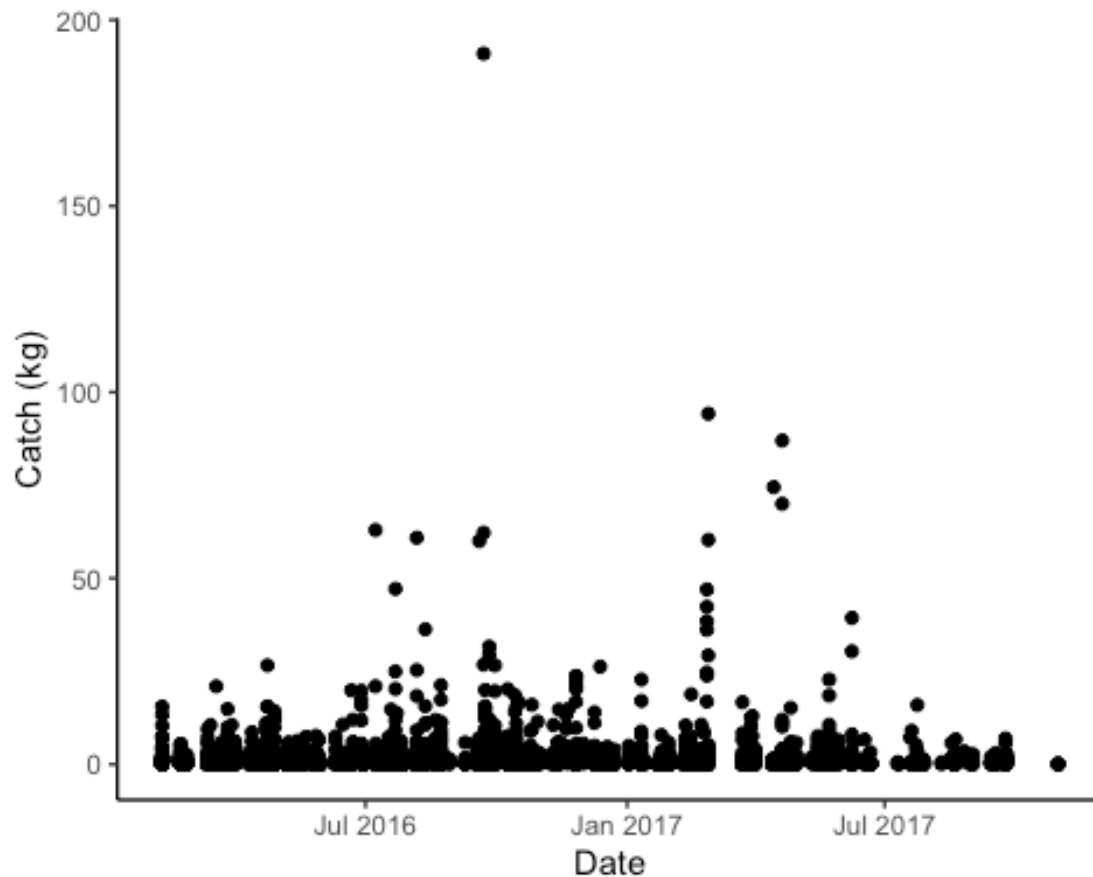
```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 4496 obs. of  15
## variables:
## $ fish_name      : chr  "laoko" "laoko" "laoko" "laoko" ...
## $ scientific_name: chr  NA NA NA NA ...
## $ family         : chr  NA NA NA NA ...
## $ genus          : chr  NA NA NA NA ...
```

```
## $ gear      : num  1 8 1 8 1 7 1 1 8 7 ...
## $ fisher_ID : num  28 9 28 9 13 24 8 8 37 25 ...
## $ date      : chr   "6/17/2017" "11/26/2016" "9/9/2016" "5/15/2017"
...
## $ month     : num   6 11 9 5 11 7 3 6 8 7 ...
## $ year      : num  2017 2016 2016 2017 2016 ...
## $ village   : num   4 2 4 3 2 4 2 2 4 4 ...
## $ time_decimal : num  2.33 4.75 2.17 2.53 9.33 2.33 3.5 3.83 3.92 1 ...
## $ total_weight : num  0.525 1.945 0.105 0.025 2.34 ...
## $ number_fish : num   1 1 2 1 2 1 1 1 2 3 ...
## $ total_sold  : num  0.525 0 0.105 0.025 0 0.23 0 0 0 0.69 ...
## $ total_consumed : num  0 1.95 0 0 2.34 ...
## - attr(*, "spec")=
## .. cols(
## ..   fish_name = col_character(),
## ..   scientific_name = col_character(),
## ..   family = col_character(),
## ..   genus = col_character(),
## ..   gear = col_double(),
## ..   fisher_ID = col_double(),
## ..   date = col_character(),
## ..   month = col_double(),
## ..   year = col_double(),
## ..   village = col_double(),
## ..   time_decimal = col_double(),
## ..   total_weight = col_double(),
## ..   number_fish = col_double(),
## ..   total_sold = col_double(),
## ..   total_consumed = col_double()
## .. )
```

2) Do a simple plot of the catch records through time.

```
dta2 = dta %>%
  mutate(date = as.Date(date,
    format = "%m/%d/%Y"))

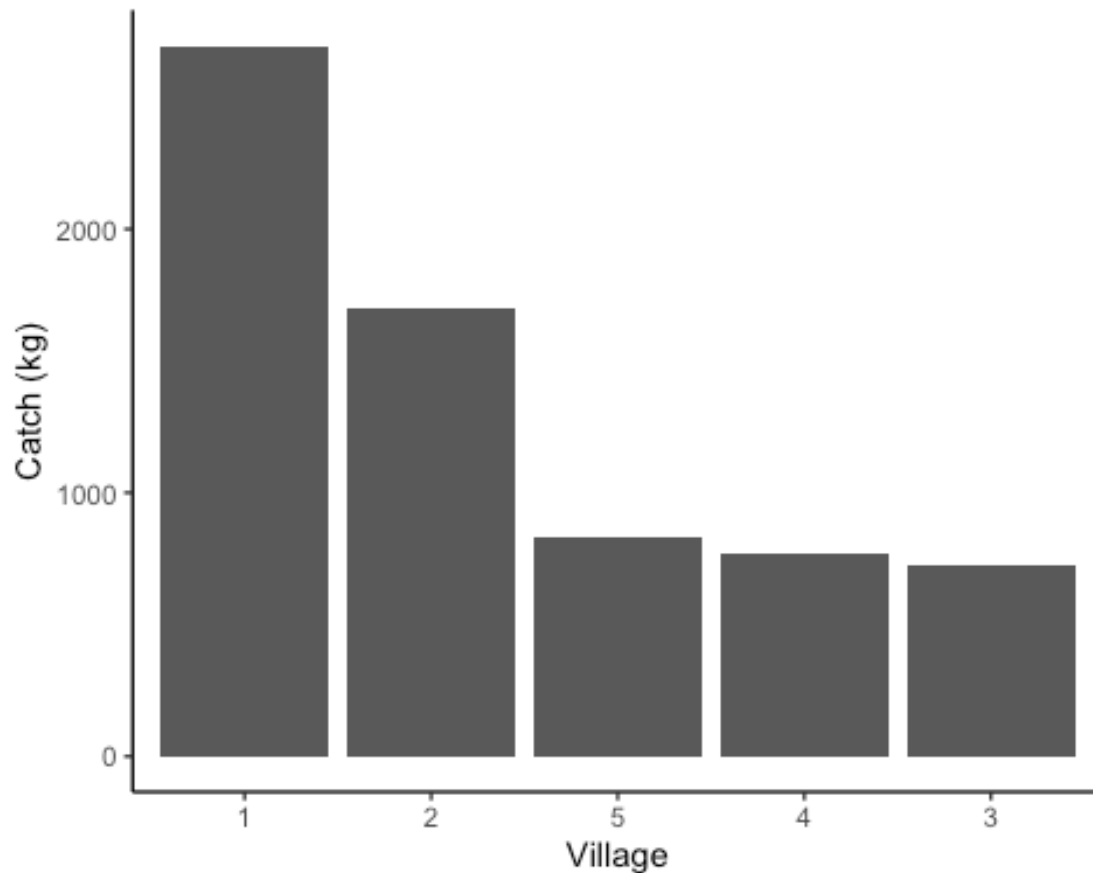
ggplot(data = dta2) +
  geom_point(aes(x = date, y = total_weight)) +
  scale_x_date(date_labels = "%b %Y") +
  labs(x = "Date", y = "Catch (kg)") +
  theme_classic()
```



3) What is the total catch per village? Create a figure to show this.

```
dta2 = dta %>%
  group_by(village) %>%
  summarise(catch = sum(total_weight))

ggplot(data = dta2) +
  geom_bar(aes(x = reorder(village, -catch), y = catch), stat = "identity") +
  labs(x = "Village", y = "Catch (kg)") +
  theme_classic()
```

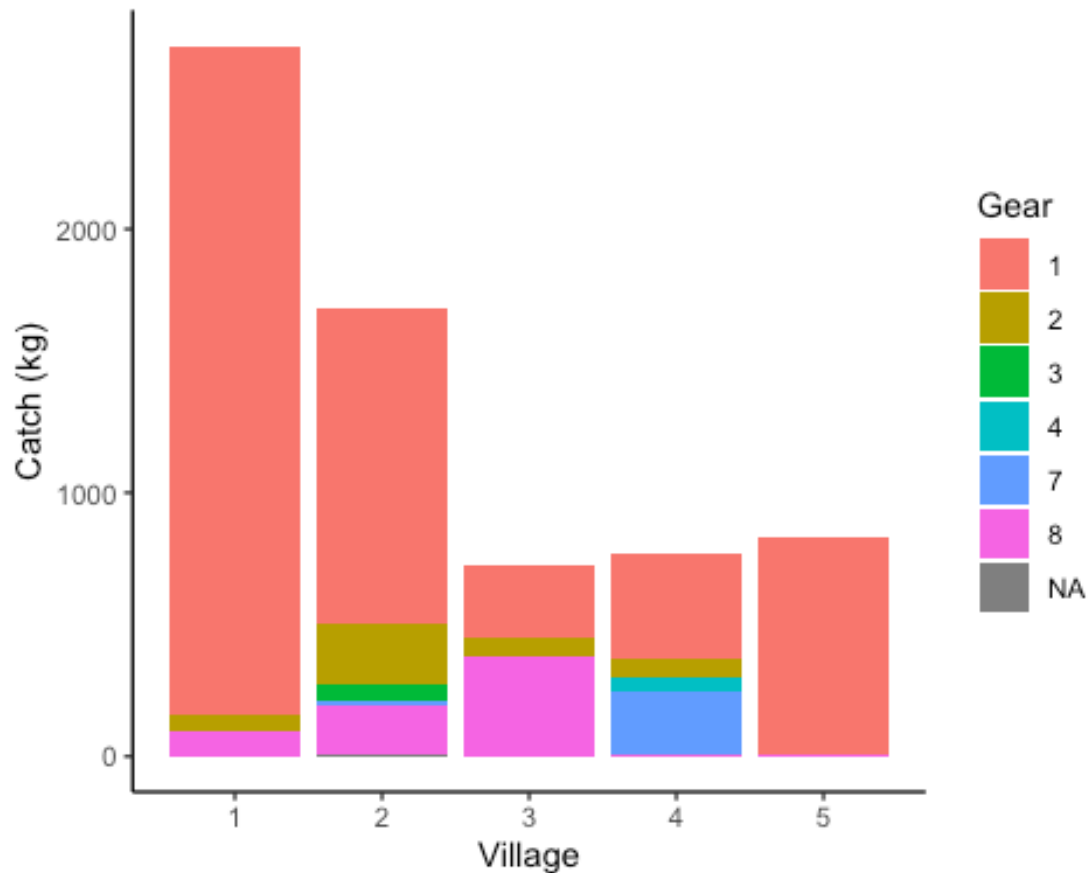


4) What is the catch per fishing gear in each village?

```
dta2 = dta %>%
  mutate(gear = as.factor(gear)) %>%
  group_by(village, gear) %>%
  summarise(catch = sum(total_weight))

## Warning: Factor `gear` contains implicit NA, consider using
## `forcats::fct_explicit_na`

ggplot(data = dta2) +
  geom_bar(aes(x = village, y = catch, fill = gear), stat = "identity") +
  labs(x = "Village", y = "Catch (kg)", fill = "Gear") +
  theme_classic()
```



5) What are the top 3 species caught in the entire bay?

```
x = dta %>%
  group_by(fish_name) %>%
  summarise(catch = sum(total_weight)) %>%
  arrange(desc(catch))

top_spp = x$fish_name[1:3]
top_spp

## [1] "tsaingatsainga" "bonique" "lagnaorana"
```

6) What is the average monthly CPUE (catch per unit effort) of the top 3 species caught in the bay? Is there any seasonality on the CPUE throughout the year? What does this mean?

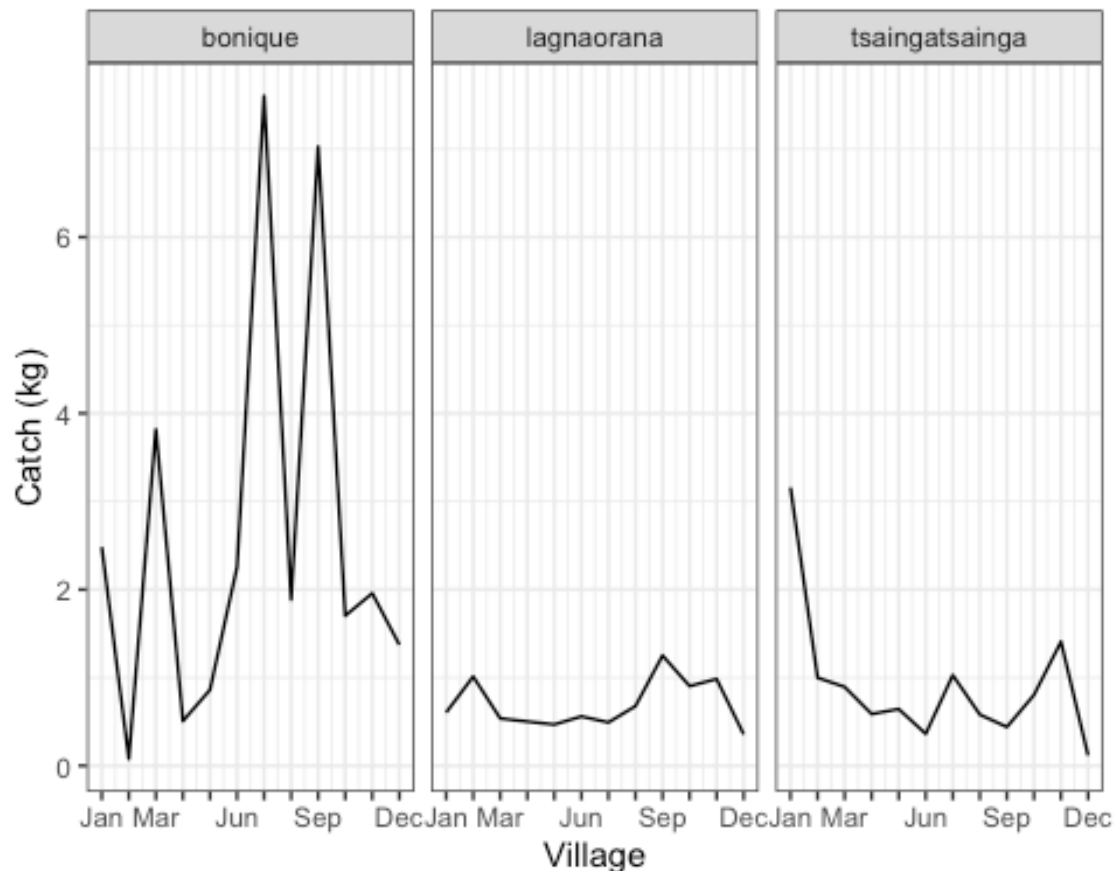
```
dta2 = dta %>%
  filter(fish_name %in% top_spp) %>%
  mutate(cpue = total_weight/time_decimal) %>%
  group_by(fish_name, month) %>%
  summarise(cpue = mean(cpue)) %>%
  ungroup()

ggplot(data = dta2) +
  geom_line(aes(x = month, y = cpue)) +
```

```

facet_wrap(~fish_name) +
  scale_x_continuous(breaks = c(1:12), labels = month.abb, guide =
guide_axis(check.overlap = T)) +
  labs(x = "Village", y = "Catch (kg)") +
  theme_bw()

```



- 7) For each village, what percentage of total catch is consumed vs sold? Try to visualize the differences between villages.

```

dta2 = dta %>%
  group_by(village) %>%
  summarise(sold = sum(total_sold),
            consumed = sum(total_consumed)) %>%
  mutate(perc_consumed = 100*consumed/(sold + consumed),
         perc_sold = 100-perc_consumed) %>%
  select(village, perc_consumed, perc_sold) %>%
  reshape2::melt(id.vars = c("village")) %>%
  mutate(variable = recode(variable,
                           "perc_sold" = "Sold",
                           "perc_consumed" = "Consumed"))

ggplot(data = dta2) +
  geom_bar(aes(x = village, y = value, fill = variable), stat = "identity") +
  labs(x = "Village", y = "%", fill = "") +

```

```
theme(axis.title = element_text(size=20),  
      axis.text = element_text(size = 20)) +  
theme_bw()
```

