Midterm_shuna

shuna

02 March, 2022

Contents


```
library(tidyverse)
library(scales)
```

Problem 1

a

```
haar_with_case_when <- function(x) {
  case_when(
    0 <= x & x < 0.5 ~ 1,
    0.5 <= x & x < 1 ~ -1,
    TRUE ~ 0
  )
}
```

b

```
haar_with_case_when(seq(-0.5, 1.5, by = 0.25))
```

```
## [1]  0  0  1  1 -1 -1  0  0  0
```

It returns the correct output.

c

```
haar_alternative <- function(x) {
  if_else(
    0 <= x & x < 0.5,
    1,
    if_else(
      0.5 <= x & x < 1,
      -1,
      0
    )
  )
}
```

d

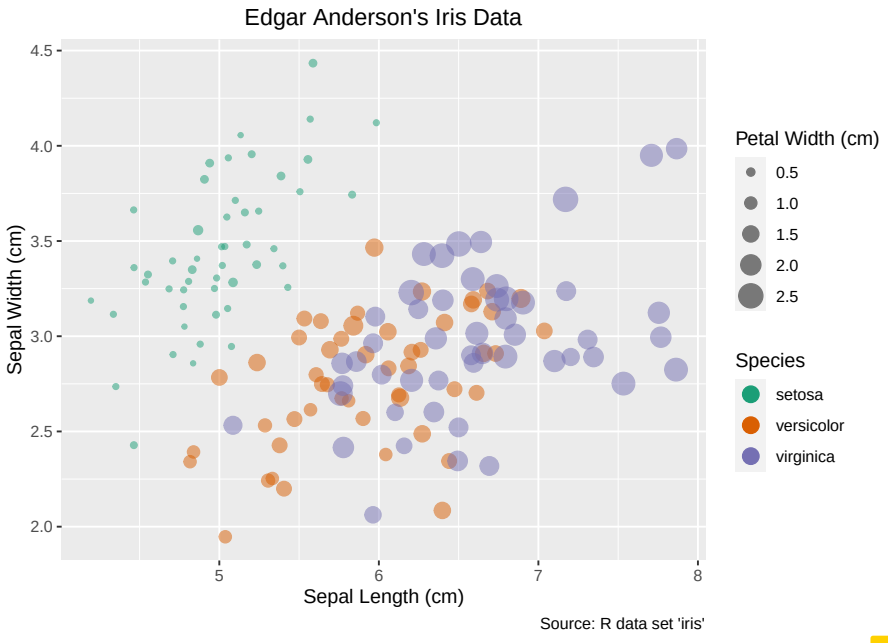
```
haar_alternative(seq(-0.5, 1.5, by = 0.25))
```

```
## [1]  0  0  1  1 -1 -1  0  0  0
```

Problem 2

```
g <- ggplot(
  data = iris,
  aes(
    x = Sepal.Length,
    y = Sepal.Width,
    size = Petal.Width,
    colour = Species
  )
) +
  geom_jitter(alpha = 0.5, height = 0.2, width = 0.2) +
  labs(
    x = "Sepal Length (cm)",
    y = "Sepal Width (cm)",
    title = "Edgar Anderson's Iris Data",
    caption = "Source: R data set 'iris'"
  ) +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_radius(name = "Petal Width (cm)") +
  scale_colour_brewer(palette = "Dark2") +
  guides(colour = guide_legend(override.aes = list(alpha = 1, size = 4)))

g
```

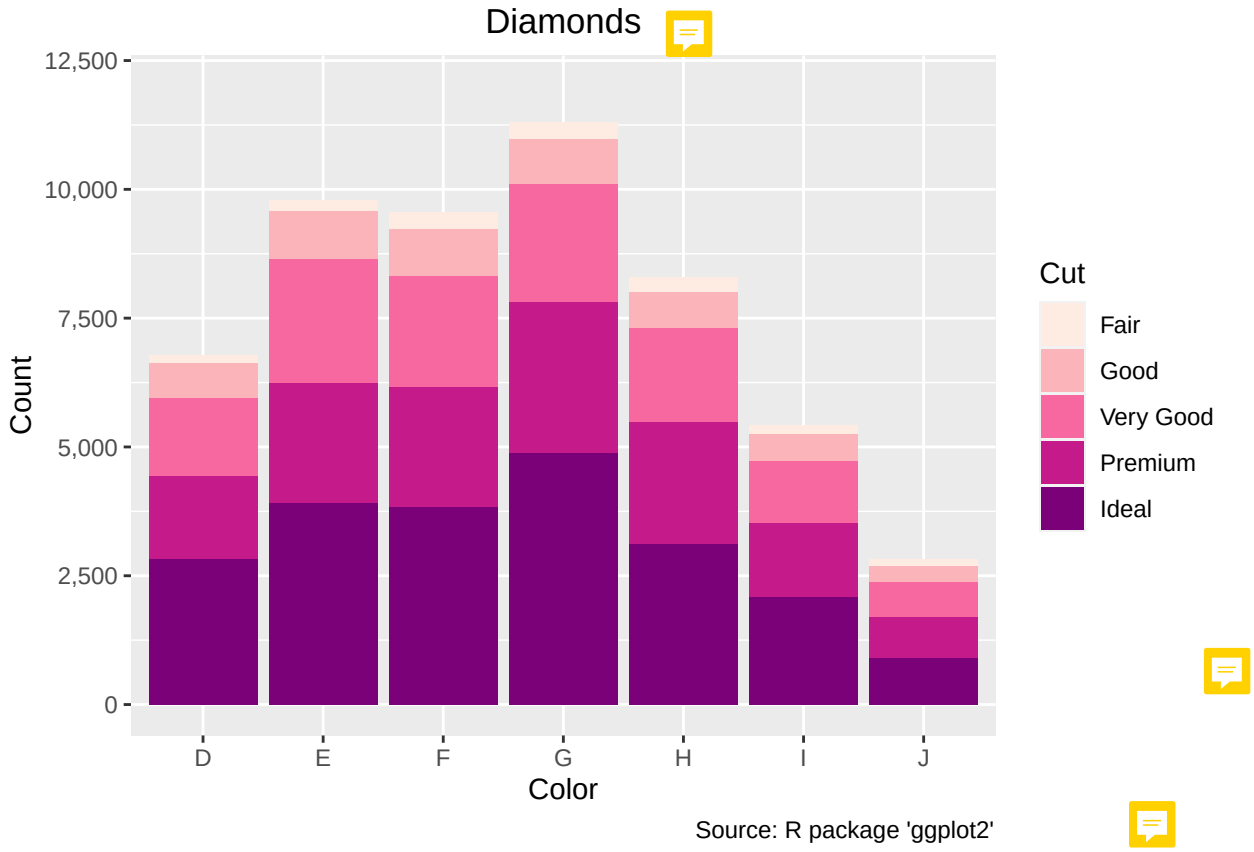


Problem 3

The plot is uninformative because both colour and cut are discrete variables. The scatter plot is used to visualize the trend in continuous variables. Every point in the scatterplot represents multiple datapoints which misdelivers the correct message.

b

```
# reference
# https://stulp.gmw.rug.nl/dataviz/twodiscretevariables.html
ggplot(diamonds, aes(color, fill = cut)) +
  geom_bar() +
  labs(
    x = "Color",
    y = "Count",
    title = "Diamonds",
    caption = "Source: R package 'ggplot2'"
  ) +
  theme(plot.title = element_text(hjust = 0.5)) +
  # reference
  # https://colorbrewer2.org/#type=sequential&scheme=RdPu&n=5
  scale_fill_brewer(
    palette = "RdPu",
    name = "Cut"
  ) +
  scale_y_continuous(labels = comma, limits = c(0, 12000))
```



Problem 4

```
unique(gss_cat$marital)
```

```
## [1] Never married Divorced      Widowed       Married       Separated
## [6] No answer
## Levels: No answer Never married Separated Divorced Widowed Married
```

a

```
gss_cat <- gss_cat |>
  filter(marital != "No answer")
# from prof's solution for the factor assignment
gss_cat <- gss_cat[!(is.na("age"))]
```

b

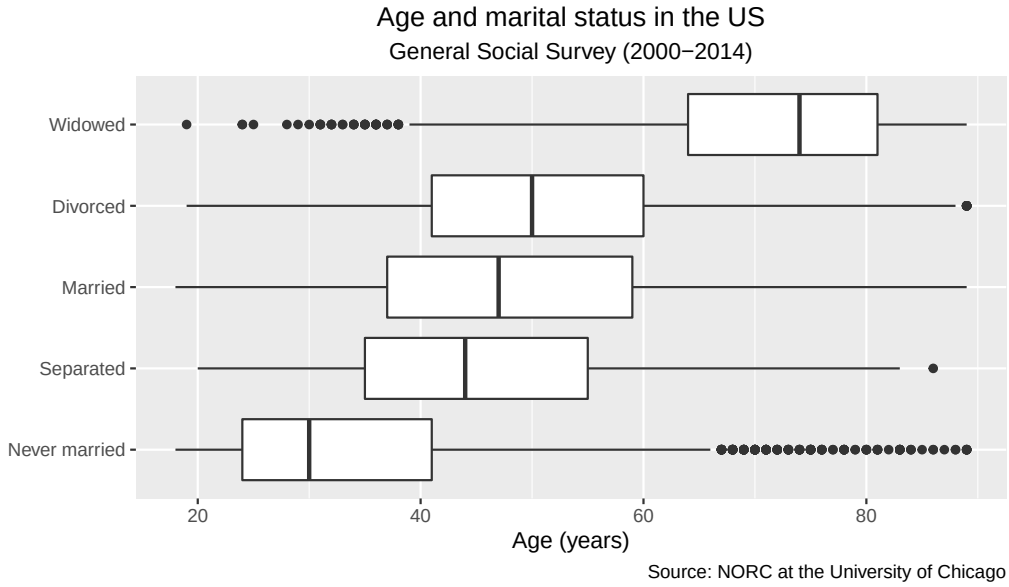
```
gss_cat$marital <- fct_drop(gss_cat[["marital"]])

# fct_reorder did not work for some reason
# tried fct_reorder(gss_cat$marital, gss_cat$age, median)
# so hardcoded
gss_cat$marital <-
  factor(
```

```
    gss_cat$marital,
    levels = c("Never married", "Separated", "Married", "Divorced", "Widowed"))
```

c

```
# reference to sorting the boxplot
# https://datavizpyr.com/reorder-boxplots-in-r-with-ggplot2/
ggplot(gss_cat, aes(x = reorder(marital, age, na.rm = TRUE), y = age)) +
  geom_boxplot(na.rm = TRUE) +
  coord_flip() +
  labs(
    y = "Age (years)",
    x = "",
    title = "Age and marital status in the US",
    subtitle = "General Social Survey (2000-2014)",
    caption = "Source: NORC at the University of Chicago"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5)
  )
```



Problem 5

```
ChickWeight_aggregate <- aggregate(
  weight ~ Time,
  data = ChickWeight,
  FUN = mean
)
```

```
chick <- as.data.frame(ChickWeight)
chick |>
  # reference
```

```r
# https://www.r-bloggers.com/2015/08/spaghetti-plots-with-ggplot2-and-ggvis/
ggplot(aes(Time, weight)) +
geom_smooth(
  data = ChickWeight_aggregate,
  method = "lm",
  aes(color = "Average\nover all\ndiets"),
) +
# reference
# https://stackoverflow.com/questions/36276240/how-to-add-legend-to-geom-smooth-in-ggplot-in-r
scale_color_manual(
  values = "green4",
  name = ""
) +
facet_wrap(~Diet) +
geom_line(alpha = 0.5, aes(group = Chick)) +
labs(
  x = "Time (days)",
  y = "Weight (g)",
  title = "Weight vs Age of Chicks on Different Diets (1-4)",
  caption = "Source: R data set 'ChickWeight'"
) +
theme(plot.title = element_text(hjust = 0.5))
```