HW 01 - Prefresher

Paul Vermette (prv25)

2025-01-29

Setup

Load packages:

```
renv::restore()
```

```
The following package(s) will be updated:

# CRAN ---------------------------------------------------------------
- MASS           [7.3-64 -> 7.3-61]
- curl           [6.2.0 -> 6.1.0]
- R.cache        [* -> 0.16.0]
- R.methodsS3    [* -> 1.8.2]
- R.oo           [* -> 1.27.0]
- R.utils        [* -> 2.12.3]
- Rcpp           [* -> 1.0.14]
- V8             [* -> 6.0.0]
- bigD           [* -> 0.3.0]
- bitops         [* -> 1.0-9]
- credentials    [* -> 2.0.2]
- desc           [* -> 1.4.3]
- gert           [* -> 2.1.4]
- gh             [* -> 1.4.1]
- gitcreds       [* -> 0.1.2]
- gt             [* -> 0.11.1]
- htmlwidgets    [* -> 1.6.4]
- httr2          [* -> 1.1.0]
- ini            [* -> 0.3.1]
- juicyjuice     [* -> 0.1.0]
- markdown       [* -> 1.13]
- reactR         [* -> 0.6.1]
- reactable      [* -> 0.4.4]
- rprojroot      [* -> 2.0.4]
- styler         [* -> 1.10.3]
- usethis        [* -> 3.1.0]
- whisker        [* -> 0.4.1]
- zip           [* -> 2.3.1]

# RSPM -----------------------------------------------------------------------
- commonmark    [* -> 1.9.2]

# Downloading packages -------------------------------------------------------
- Downloading curl from CRAN ...                OK [file is up to date]
- Downloading MASS from CRAN ...                OK [file is up to date]
- Downloading R.cache from CRAN ...             OK [file is up to date]
- Downloading R.methodsS3 from CRAN ...         OK [file is up to date]
- Downloading R.oo from CRAN ...                OK [file is up to date]
- Downloading R.utils from CRAN ...             OK [file is up to date]
- Downloading Rcpp from CRAN ...                OK [file is up to date]
- Downloading V8 from CRAN ...                  OK [file is up to date]
- Downloading bigD from CRAN ...                OK [file is up to date]
- Downloading bitops from CRAN ...              OK [file is up to date]
- Downloading commonmark from RSPM ...          OK [file is up to date]
- Downloading credentials from CRAN ...         OK [file is up to date]
- Downloading desc from CRAN ...                OK [file is up to date]
- Downloading gert from CRAN ...                OK [file is up to date]
- Downloading zip from CRAN ...                 OK [file is up to date]
- Downloading gh from CRAN ...                  OK [file is up to date]
- Downloading gitcreds from CRAN ...            OK [file is up to date]
- Downloading httr2 from CRAN ...               OK [file is up to date]
- Downloading ini from CRAN ...                 OK [file is up to date]
- Downloading gt from CRAN ...                  OK [file is up to date]
- Downloading htmlwidgets from CRAN ...         OK [file is up to date]
- Downloading juicyjuice from CRAN ...          OK [file is up to date]
- Downloading markdown from CRAN ...            OK [file is up to date]
- Downloading reactable from CRAN ...           OK [file is up to date]
- Downloading reactR from CRAN ...              OK [file is up to date]
- Downloading rprojroot from CRAN ...           OK [file is up to date]
- Downloading styler from CRAN ...              OK [file is up to date]
- Downloading usethis from CRAN ...             OK [file is up to date]
- Downloading whisker from CRAN ...             OK [65.3 Kb in 0.35s]
Successfully downloaded 29 packages in 11 seconds.

# Installing packages --------------------------------------------------------
- Installing curl ...                           OK [installed binary and cached
in 0.67s]
- Installing MASS ...                           OK [installed binary and cached
in 0.66s]
- Installing R.methodsS3 ...                    OK [installed binary and cached
in 0.56s]
- Installing R.oo ...                           OK [installed binary and cached
in 0.6s]
- Installing R.utils ...                        OK [installed binary and cached
in 0.71s]
```

```
- Installing R.cache ...                        OK [installed binary and cached
in 0.59s]
- Installing Rcpp ...                           OK [installed binary and cached
in 2.7s]
- Installing V8 ...                             OK [installed binary and cached
in 1.7s]
- Installing bigD ...                           OK [installed binary and cached
in 0.56s]
- Installing bitops ...                         OK [installed binary and cached
in 0.56s]
- Installing commonmark ...                     OK [installed binary and cached
in 0.68s]
- Installing credentials ...                    OK [installed binary and cached
in 1.1s]
- Installing desc ...                           OK [installed binary and cached
in 0.87s]
- Installing zip ...                            OK [installed binary and cached
in 0.93s]
- Installing gert ...                           OK [installed binary and cached
in 1.1s]
- Installing gitcreds ...                       OK [installed binary and cached
in 0.55s]
- Installing httr2 ...                          OK [installed binary and cached
in 0.75s]
- Installing ini ...                            OK [installed binary and cached
in 0.52s]
- Installing gh ...                             OK [installed binary and cached
in 0.88s]
- Installing htmlwidgets ...                    OK [installed binary and cached
in 0.79s]
- Installing juicyjuice ...                     OK [installed binary and cached
in 0.54s]
- Installing markdown ...                       OK [installed binary and cached
in 0.57s]
- Installing reactR ...                         OK [installed binary and cached
in 0.87s]
- Installing reactable ...                      OK [installed binary and cached
in 0.89s]
- Installing gt ...                             OK [installed binary and cached
in 1.4s]
- Installing rprojroot ...                      OK [installed binary and cached
in 0.62s]
- Installing styler ...                         OK [installed binary and cached
in 1.2s]
- Installing whisker ...                        OK [installed binary and cached
in 0.57s]
- Installing usethis ...                        OK [installed binary and cached
in 0.94s]
```

```
library(tidyverse)
library(googlesheets4)
library(janitor)
library(scales)
```

Exercise 1

```
# Authenticate Google Sheets
gs4_auth()

# Import the data
sheet_url <- "https://docs.google.com/spreadsheets/d/1Rxz5IFE16bA9pPAY5_B5
LhitvPDIlYrlMwp7fmszwOA/edit?gid=0#gid=0"
data_raw <- read_sheet(sheet_url, col_types = "c")
colnames(data)
```

```
NULL
```

```
data_clean <- data_raw %>%
  clean_names()

colnames(data_clean)
```

```
 [1] "year"                       "season"
 [3] "season_order"               "title"
 [5] "author"                     "publisher"
 [7] "date"                       "description"
 [9] "isbn"                       "github_permalink"
[11] "google_thumbnail"           "google_small_thumbnail"
[13] "cover_url_override"         "cover_url"
[15] "github_thumbnail"           "cover"
[17] "style"                      "man_partially_unclothed"
[19] "woman_partially_unclothed" "has_poc"
[21] "x21"
```

```
data_clean <- data_clean %>%
  select(year,
         title,
         author,
         publisher,
         date,
         description,
         style,
         "man_partially_unclothed",
         "woman_partially_unclothed",
         "has_poc") %>%
  mutate( date = as.character(date),
    date = str_trim(date),
    date = str_remove_all(date, "[^0-9/-]"),
    date = if_else( str_detect( date, "/"),
      suppressWarnings( as.character( mdy( date))),
      suppressWarnings( as.character( ymd( date)))
    ),

    date = if_else(
      is.na( date ),
      suppressWarnings( as.character( dmy( date))),
      date
    )
  )

glimpse(data_clean)
```

```
Rows: 1,435
Columns: 10
$ year                      <chr> "2011", "2011", "2011", "2011", "2011", "201…
$ title                     <chr> "Too Rich for a Bride", "Breaking the Rules"…
$ author                    <chr> "Mona Hodgson", "Suzanne Brockmann", "Mary B…
$ publisher                 <chr> "Waterbrook", "Ballantine", "Delacorte", "Gr…
$ date                      <chr> "2011-05-01", "2011-04-01", "2011-07-03", "2…
$ description               <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "Par…
$ style                     <chr> "Photorealistic", "Photorealistic", "Photore…
$ man_partially_unclothed   <chr> "FALSE", "FALSE", "FALSE", "FALSE", "FALSE",…
$ woman_partially_unclothed <chr> "FALSE", "FALSE", "FALSE", "FALSE", "FALSE",…
$ has_poc                   <chr> "FALSE", "FALSE", "FALSE", "FALSE", "TRUE", …
```

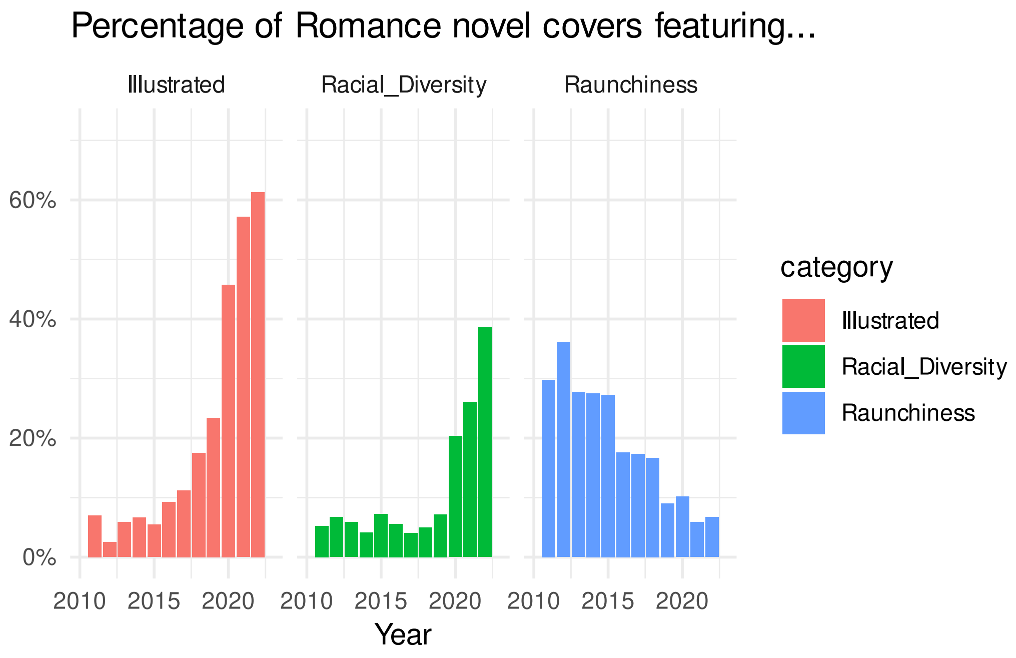
Exercise 2

```
library(dplyr)
library(tidyr)

annual_percentages <- data_clean %>%
  group_by(year) %>%
  summarise(
              Raunchiness    =    mean((man_partially_unclothed    ==    TRUE    |
woman_partially_unclothed == TRUE), na.rm = TRUE) * 100,
    Illustrated = mean(style == "Illustrated", na.rm = TRUE) * 100,
    Racial_Diversity = mean(has_poc == TRUE, na.rm = TRUE) * 100
  ) %>%
  pivot_longer(cols = c(Raunchiness, Illustrated, Racial_Diversity),
```

```r
                names_to = "category",
                values_to = "percentage")

ggplot(annual_percentages, aes(x = as.numeric(year), y = percentage, fill =
category)) +
  geom_col() +
  facet_wrap(~ category) +
  labs(
    title = "Percentage of Romance novel covers featuring...",
    x = "Year",
    y = element_blank()
  ) +
  scale_x_continuous(
   breaks = seq(2010, max(as.numeric(annual_percentages$year)), by = 5), # Show
every 5 years
    limits = c(2010, max(as.numeric(annual_percentages$year))) # Ensure 2010 is
included
  )  +
  scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  theme_minimal()
```



Exercise 3

```r
data_eating_animals <- read.csv("data/eating-animals.csv")
```

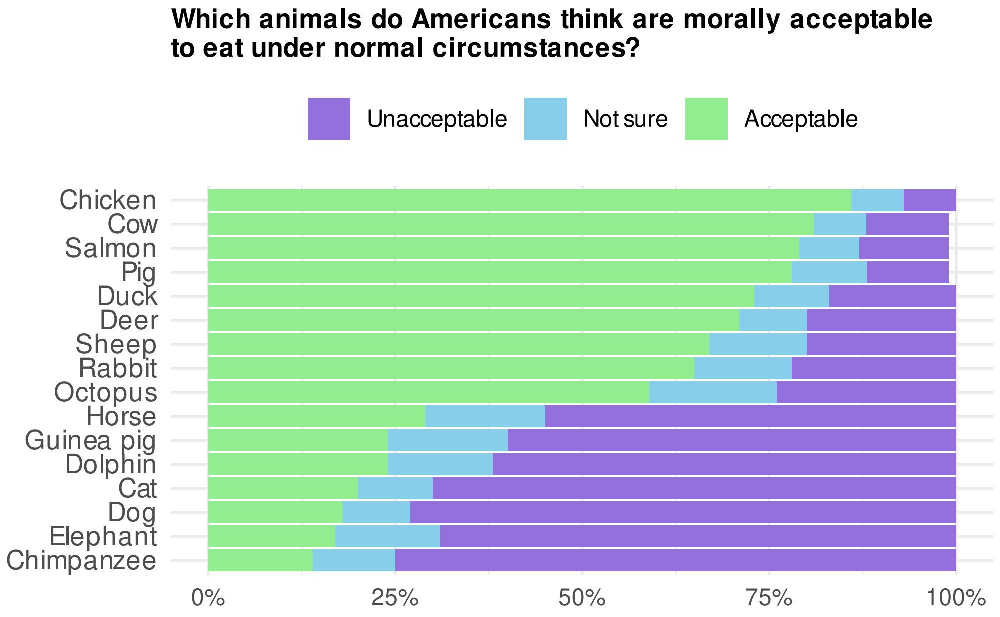
```r
data_eating_animals2 <- data_eating_animals %>%
    rename("Not sure" = Not.sure) %>%
   mutate(
    Acceptable_num = as.numeric(sub("%", "", Acceptable)),
    Animal = fct_reorder(Animal, Acceptable_num)) %>%

  pivot_longer(cols = c(Acceptable, Unacceptable, `Not sure`),
              names_to = "response",
              values_to = "percentage") %>%

# Reversing the order
  mutate(
    percentage = as.numeric(sub("%", "", percentage)),
    response = fct_rev(factor(response, levels = c("Acceptable", "Not sure",
"Unacceptable"), ordered = TRUE)
  )
)
ggplot(data_eating_animals2,
       aes(x = percentage,
           y = Animal,
           fill = response)) +

  geom_col() +
  labs(
    title = "Which animals do Americans think are morally acceptable\nto eat
under normal circumstances?",
    x = "",
    y = ""
  ) +
  scale_fill_manual(
    name = NULL,
    values = c(
      "Acceptable" = "#90EE90",
      "Not sure" = "#87CEEB",
      "Unacceptable" = "#9370DB"


    )
  ) +
  scale_x_continuous(labels = scales::percent_format(scale = 1)) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 10, face = "bold"),
    axis.text.y = element_text(size = 10),
    legend.position = "top",
    plot.caption = element_text(hjust = 1, size = 9)
  )
```



Exercise 4

```
commissary_data <- read.csv("data/commissary-prices.csv")

commissary_data <- commissary_data %>%
  mutate(price = as.numeric(str_remove(price, "\\$"))) # used double slash
because $ is used in R already.

ramen_prices <- commissary_data %>%
  filter(str_detect(tolower(product_type), "ramen")) %>%
  group_by(state) %>%
  summarize(avg_price = mean(price, na.rm = TRUE)) %>%
  arrange(desc(avg_price)) %>%
  slice_head(n = 10)

ramen_analysis <- commissary_data %>%
  filter(str_detect(tolower(product_type), "ramen")) %>%
  mutate(price_clean = as.numeric(str_remove(price, "\\$"))) %>%
  group_by(state) %>%
  summarize(
    average_price = mean(price_clean, na.rm = TRUE)
  ) %>%

  arrange(desc(average_price)) %>%
  slice_head(n = 10)

print(ramen_analysis)
```

```
# A tibble: 10 × 2
   state       average_price
   <chr>               <dbl>
 1 Florida             1.06
 2 Hawaii              0.895
 3 Mississippi         0.89
 4 Maine               0.8
 5 Vermont             0.8
 6 Kansas              0.72
 7 Georgia             0.7
 8 Missouri            0.7
 9 New Mexico          0.69
10 Kentucky            0.68
```

Exercise 5

```
deodorant_analysis <- commissary_data %>%
  filter(str_detect(tolower(product_type), "deodorant")) %>%
  mutate(price_clean = as.numeric(str_remove(price, "\\$"))) %>%
  group_by(state) %>%
  summarize(
    min_price = min(price_clean, na.rm = TRUE)
  ) %>%
  arrange(min_price) %>%
  slice_head(n = 10)

print(deodorant_analysis)
```

```
# A tibble: 10 × 2
   state          min_price
   <chr>              <dbl>
 1 Ohio                0.06
 2 Michigan            0.1
 3 Georgia             0.25
 4 Arizona             0.4
 5 Illinois            0.44
 6 Pennsylvania        0.52
 7 Maine               0.58
 8 Kentucky            0.65
 9 Missouri            0.68
10 North Carolina      0.7
```

Exercise 6

```
lady_speedstick_count <- commissary_data %>%
  filter(str_detect(tolower(description), "lady speed stick")) %>%
  distinct(state) %>%
  nrow()

cat("Number of states selling Lady Speed Stick products:", lady_speedstick_count,
"\n")
```

```
Number of states selling Lady Speed Stick products: 14
```

GAI self-reflection

I used GAI on Claude, with its ‘exploratory’ section to tell me how to think about my code. I am a new data student, so a lot is unknown to me. The greatest challenge I recognized after consistent failure with visual display was that I did not consider how to properly put the sections in the proper order. I couldn’t figure out the legend for the death of me. The last 3 sections were relatively easy compared to 2 & 3. I absolutely need to practice more on visualizations in R.