Extra credit

INFO 2950 - Spring 2023

Lili Mkrtchyan

4/11/01

Setup

Load packages and data:

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.0
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(stringr)
library(survival)

survivalists <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytue
```

```
Rows: 94 Columns: 16
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (10): name, gender, city, state, country, reason_tapped_out, reason_cate...
dbl  (5): season, age, result, days_lasted, day_linked_up
lgl  (1): medically_evacuated
```

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
loadouts <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday
```

```
Rows: 940 Columns: 6
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): version, name, item_detailed, item
dbl (2): season, item_number

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
episodes <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday
```

```
Rows: 98 Columns: 11
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (4): version, title, quote, author
dbl  (6): season, episode_number_overall, episode, viewers, imdb_rating, n_r...
date (1): air_date

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
seasons <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/
```

```
Rows: 9 Columns: 8
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (3): version, location, country
dbl  (4): season, n_survivors, lat, lon
date (1): date_drop_off

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Extra credit

Research questions:

Do loadouts have a long term effect on the determination of the winner?

How safe is Survivor? How many people get medically evacuate?

What factors play role in winning/succeeding in the show? Gender, profession, loadouts, age?

```
library(viridis)
```

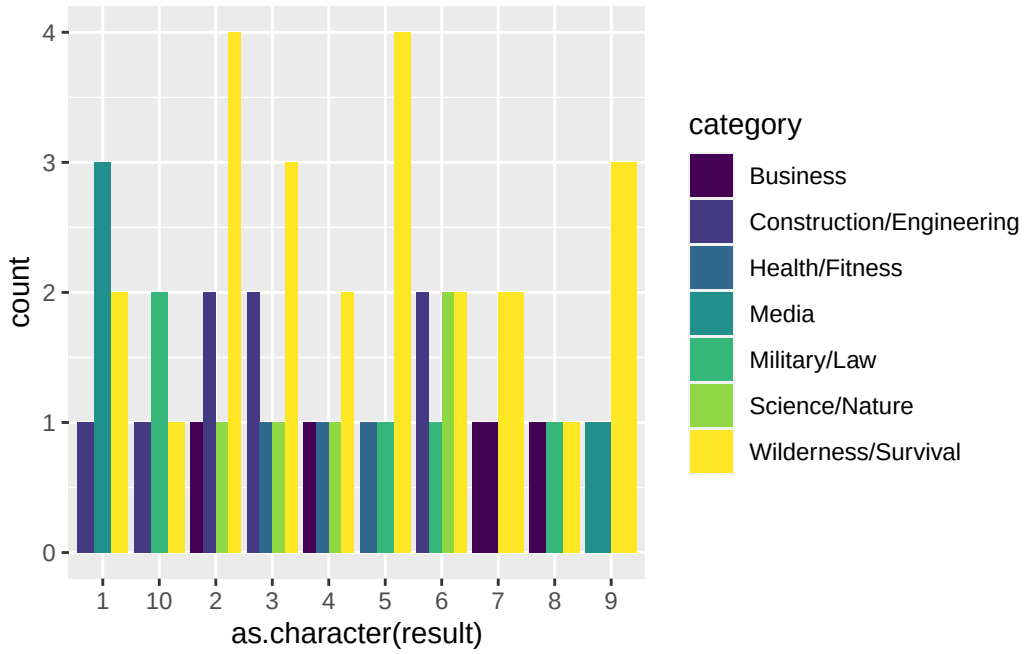
```
Loading required package: viridisLite
```

```
survivalists_clean <- survivalists |>
  select(season, name, age, gender, result,
         days_lasted, medically_evacuated,
         reason_category, profession)

exis_professions <- survivalists_clean |>
  select(profession) |>
  group_by(profession) |>
  summarize()

# Diving the professions into categories
wilderness_survival <- exis_professions |>
  filter(str_detect(profession, 'Survival|Wild|Primitive|Skil|Outdoor')) |>
  mutate(category = 'Wilderness/Survival')

business <- exis_professions |>
  filter(str_detect(profession, 'Manager|Accountant|Commercial')) |>
  mutate(category = 'Business')

construction_eng <- exis_professions |>
  filter(str_detect(profession, 'Engineer|Construct|Electr|Build|Operat|Make')) |>
  mutate(category = 'Construction/Engineering')

sci_nature <- exis_professions |>
  filter(str_detect(profession, 'Nature|Sci|Bio|Env|Agr|Anthro')) |>
  mutate(category = 'Science/Nature')
```

```r
military_law <- exis_professions |>
  filter(str_detect(profession, 'US|Army|Navy|Enforcement')) |>
  mutate(category ='Military/Law')

health <- exis_professions |>
  filter(str_detect(profession, 'Herbalist|Physician|Psychotherapist')) |>
  mutate(category = 'Health/Fitness')

media <- exis_professions |>
  filter(str_detect(profession, 'Photo|Writer|Media|Video|Author')) |>
  mutate(category = 'Media')

# Creating a unified dataframe with the professions and respective categories
prof_categories <- rbind(wilderness_survival,
                         business, construction_eng, sci_nature,
                         military_law, health, media)

# Delete the miscategorized rows
prof_categories <- prof_categories[-31,]

# Add professions that were left out from defined categories
all_prof_categories <- merge(exis_professions, prof_categories, all.x = TRUE)
all_prof_categories[is.na(all_prof_categories)] <- 'Other'

# Merge the two data frames to include profession categories
survivalists_clean <- merge(survivalists_clean, prof_categories, by = "profession")

# Merge original data frame to include the number of participants that have a certain prof
temp_graph <- survivalists_clean |>
  select(result, category) |>
  group_by(category, result) |>
  summarize(count = n())
```

```
`summarise()` has grouped output by 'category'. You can override using the
`.groups` argument.
```

```r
ggplot(temp_graph, aes(x = as.character(result), y = count, fill = category)) +
  geom_col(position = "dodge") +
  scale_fill_viridis(discrete = TRUE, option = "D")
```




```
#ggplot(survivalists_clean, mapping = aes(x = days_lasted, y = count, colour = gender, fil
 # geom_line()
```