Wrangling Data

Ben Goldstone

10/6/2023

Import Dataset

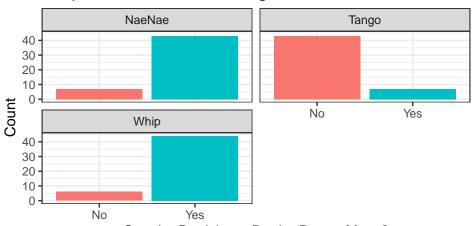
```
DanceMoves = read_csv("~/CSVs/DanceMoves.csv")
## Rows: 50 Columns: 16
## -- Column specification -----
## Delimiter: ","
## chr (14): Participant, Tango, Renegade, Whip, NaeNae, Spongebob, Woah, Floss...
## dbl (2): Flexible, Total
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
colnames (DanceMoves)
                                                  "Whip"
                                                                "NaeNae"
    [1] "Participant"
                     "Tango"
                                    "Renegade"
##
   [6] "Spongebob"
                                    "Floss"
                                                  "Moonwalk"
                                                                "Worm"
                      "Woah"
## [11] "Wop"
                                                  "Flexible"
                      "Sport"
                                    "OrgDance"
                                                                "Enjoy"
## [16] "Total"
DanceMoves2 = gather(DanceMoves, key="Move", value="CanDoDanceMove", 2:13)
DanceMoves2
## # A tibble: 600 x 6
##
     Participant Flexible Enjoy Total Move CanDoDanceMove
##
                    <dbl> <chr> <dbl> <chr> <chr>
##
                        5 No
                                    4 Tango No
  1 Person1
   2 Person2
                        8 Yes
                                   10 Tango No
## 3 Person3
                        7 Yes
                                    9 Tango No
## 4 Person4
                       10 No
                                   12 Tango Yes
## 5 Person5
                                    6 Tango No
                        9 Yes
## 6 Person6
                        1 No
                                    1 Tango No
## 7 Person7
                        6 Yes
                                    3 Tango No
## 8 Person8
                        1 No
                                    3 Tango No
## 9 Person9
                        8 Yes
                                    7 Tango No
## 10 Person10
                        4 Yes
                                    5 Tango No
## # ... with 590 more rows
# Calculate counts by Move and CanDoDanceMove
counts <- DanceMoves2 %>%
  count(Move, CanDoDanceMove)
# Filter data to include counts above a threshold (e.g., 30)
```

```
filtered_counts = counts %>%
    filter(n > 40)

# Filter the original data based on the filtered counts
DanceMoves2_filtered <- DanceMoves2 %>%
    filter(Move %in% filtered_counts$Move)

gf_bar(~CanDoDanceMove, data=DanceMoves2_filtered, fill=~CanDoDanceMove) %>%
    gf_facet_wrap(~Move, ncol=2) %>% gf_labs(title="Many Muhlenberg Students can do the NaeNae and Whip but are not able to Tango", x="Can the Participant Do the Dance Move?",
    y="Count") %>% gf_theme(legend.position="none")
```

Many Muhlenberg Students can do the NaeNae and Whip but are not able to Tango



Can the Participant Do the Dance Move?