One-Factor ANOVA

Howard Nguyen

2022-09-29

INSTALL AND LOAD PACKAGES

```r
# Install pacman ("package manager") if needed
if (!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```r
# pacman must already be installed; then load contributed
# packages (including pacman) with pacman
pacman::p_load(magrittr, pacman, psych, rio, tidyverse)
# magrittr: for pipes
# pacman: for loading/unloading packages
# psych: for descriptive statistics
# rio: for importing data
# tidyverse: for so many reasons
```

LOAD AND PREPARE DATA

```r
# Save data to "df" (for "data frame")

df <- import("../data/StateData.xlsx") %>%
  as_tibble() %>%
  select(
    state_code,
    psychRegions,
    instagram:modernDance
  ) %>%
  mutate(
    psychRegions = as.factor(psychRegions),
    psychRegions = fct_recode(psychRegions,
      "Friendly" = "Friendly and Conventional",
      "Relaxed" = "Relaxed and Creative",
      "Temperamental" = "Temperamental and Uninhibited"
    )
  ) %>%
  print()
```

```
## # A tibble: 48 x 14
##    state_code psychRegi~1 insta~2 faceb~3 retweet entre~4   gdpr privacy unive~5
##    <chr>      <fct>         <dbl>   <dbl>   <dbl>   <dbl>  <dbl>   <dbl>   <dbl>
##  1 AL         Friendly       0.64    1.65    0.35   0.257 -0.769   0.583    1.74
##  2 AZ         Relaxed       0.183  -0.259  -0.566   0.562 -0.306  -0.452  -0.771
##  3 AR         Friendly      0.456    1.10  -0.598   0.245 -0.595   0.689    0.024
##  4 CA         Relaxed        1.47  -0.422   0.481   0.502   1.12   0.231   -1.92
##  5 CO         Friendly      -1.03   -1.06  -0.902   0.023  0.588  -0.215  -0.444
##  6 CT         Temperamen~   0.374  -0.982    1.14   0.069  0.712   0.362    0.37
##  7 DE         Temperamen~    1.48   -1.12    1.19    2.55   1.21   0.904    2.19
##  8 FL         Friendly       0.85    0.38   -0.23   0.783 -0.231  -0.137   -1.12
##  9 GA         Friendly      0.807   0.526   0.035    1.95  0.403  -0.398   -0.24
## 10 ID         Relaxed      -0.736  -0.269   -1.80   0.296 -0.174   0.075  -0.044
## # ... with 38 more rows, 5 more variables: mortgage <dbl>, volunteering <dbl>,
## #   museum <dbl>, scrapbook <dbl>, modernDance <dbl>, and abbreviated variable
## #   names 1: psychRegions, 2: instagram, 3: facebook, 4: entrepreneur,
## #   5: university
```

EXPLORE DATA

```
# Bar chart of group frequencies
df %>%
  ggplot() +
  geom_bar(
    aes(
      x    = psychRegions,  # Variable to chart
      fill = psychRegions   # Color bars by variable
    )
  ) +
  theme(legend.position = "none")
```
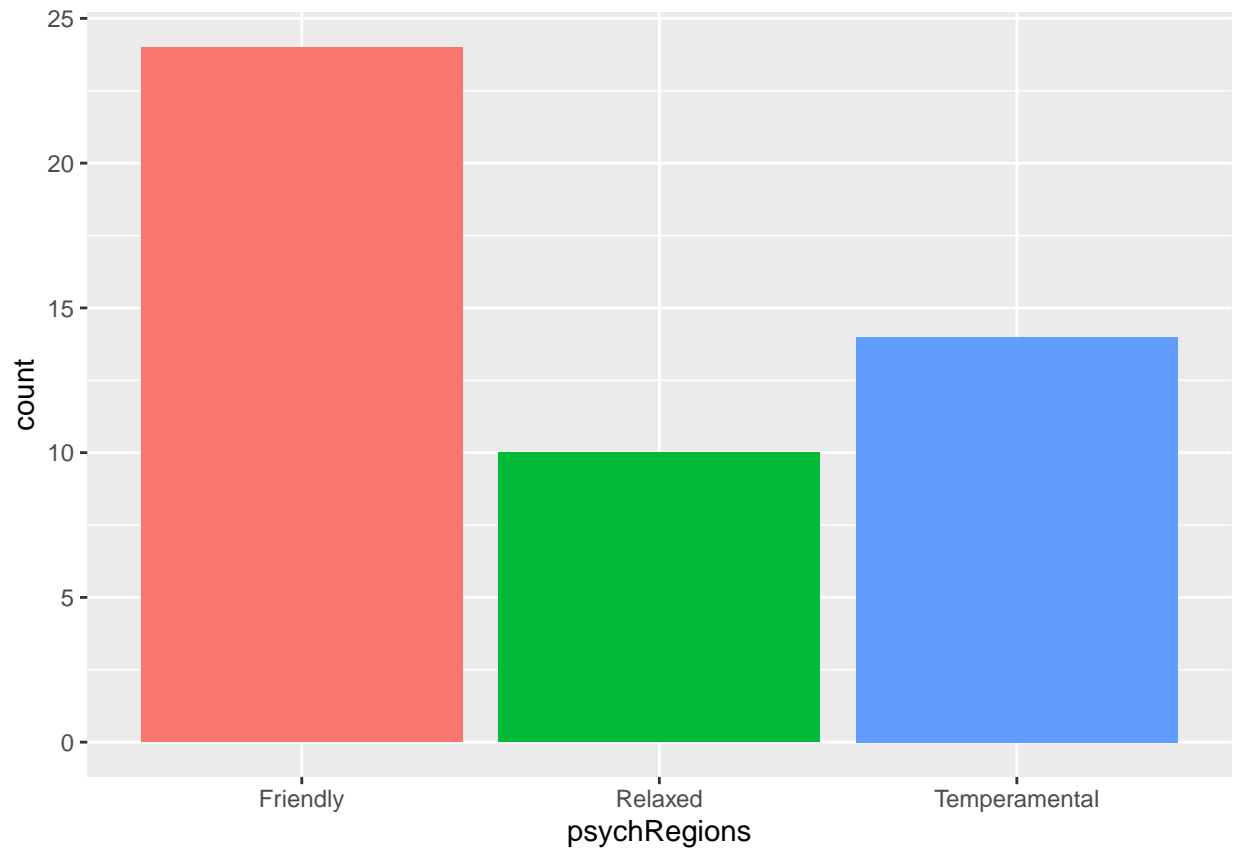
```r
# Boxplots
df %>%
  ggplot(
    aes(
      x    = psychRegions,  # Grouping variable
      y    = volunteering,  # Outcome variable
      fill = psychRegions   # Color variable
    )
  ) +
  geom_boxplot() +
  coord_flip() +
  xlab("") +
  theme(legend.position = "none")
```
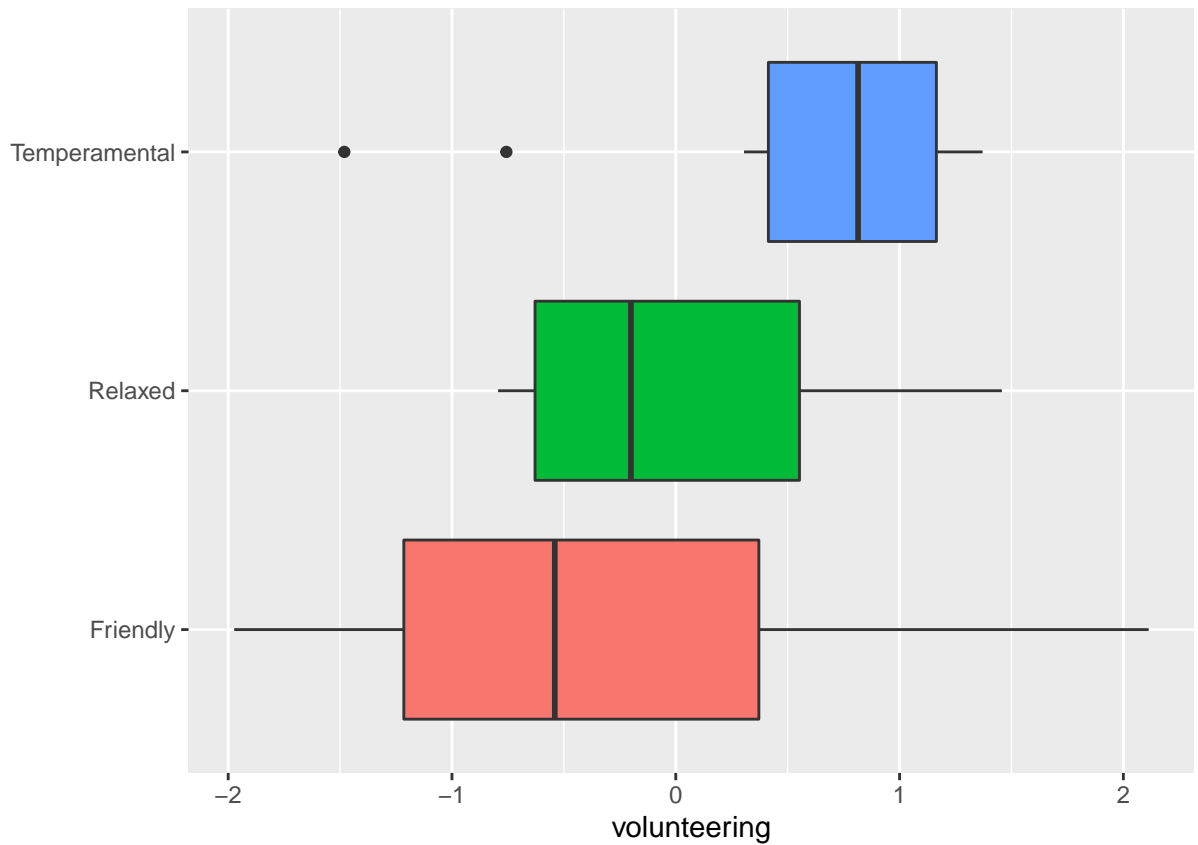
```
# Density plots
df %>%
  ggplot(
    aes(
      x    = volunteering,  # Outcome variable
      fill = psychRegions    # Color/grouping variable
    )
  ) +
  geom_density(alpha = 0.5) +
  theme(legend.position = "bottom")
```
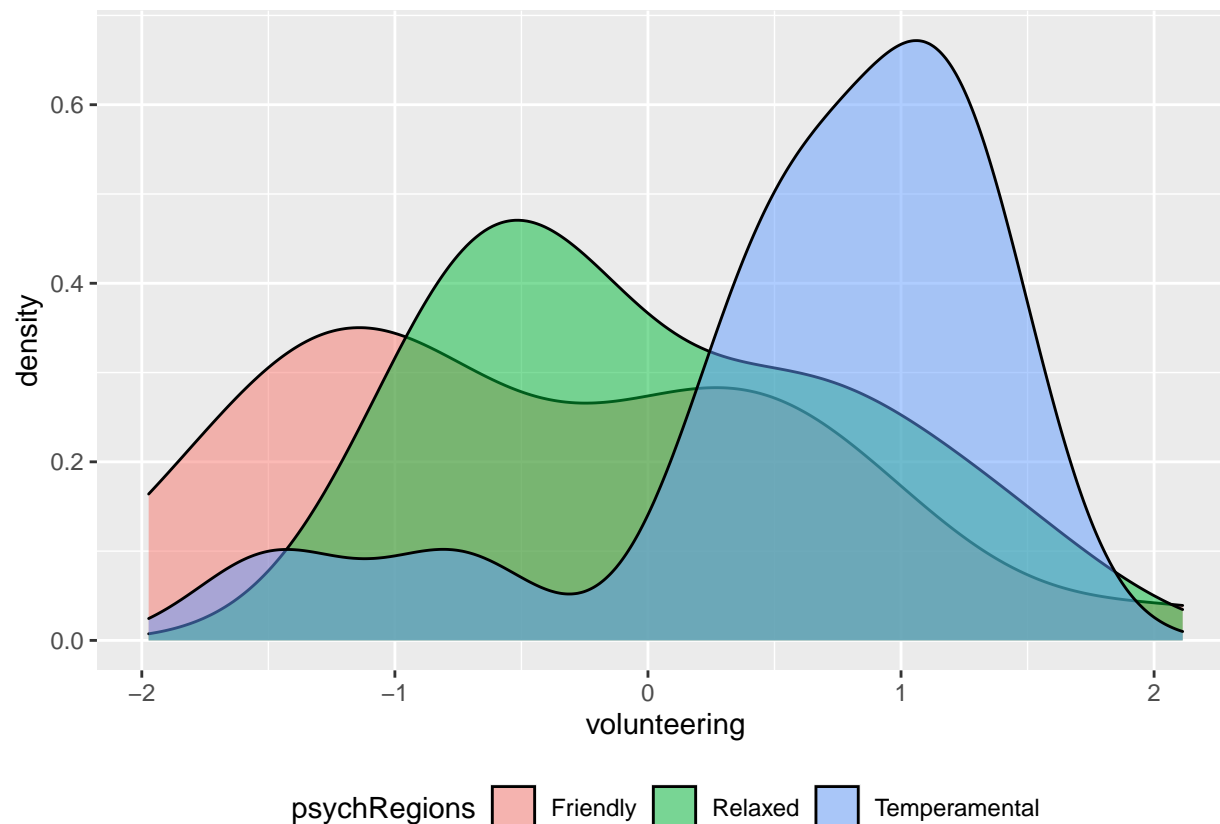
```r
# Descriptive statistics by group
df %$%               # Exposition pipe
  describeBy(        # describeBy function from psych
    volunteering,    # Outcome variable
    psychRegions     # Grouping variable
  )
```

```
## 
##  Descriptive statistics by group
## group: Friendly
##    vars  n mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 24 -0.4 1.04  -0.54   -0.44 1.15 -1.97 2.11  4.09 0.46    -0.67 0.21
## ----------------------------------------------------------------
## group: Relaxed
##    vars  n mean   sd median trimmed  mad   min  max range skew kurtosis   se
## X1    1 10 0.03 0.79   -0.2   -0.04 0.86 -0.79 1.46  2.25 0.46    -1.43 0.25
## ----------------------------------------------------------------
## group: Temperamental
##    vars  n mean   sd median trimmed mad   min  max range  skew kurtosis   se
## X1    1 14 0.61 0.82   0.81    0.72 0.6 -1.48 1.37  2.85 -1.31     0.69 0.22
```

ONE-WAY ANOVA

```
# Conduct one-way ANOVA
fit <- df %>%
  aov(
    volunteering ~ psychRegions,  # "as a function of"
    data = .
  )
```

```
# Show ANOVA table
fit %>% summary()
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## psychRegions   2   8.97   4.486   5.154 0.00965 **
## Residuals     45  39.16   0.870
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

POST-HOC TESTS

```
# TukeyHSD has its own function
fit %>% TukeyHSD()
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = volunteering ~ psychRegions, data = .)
##
## $psychRegions
##                            diff        lwr      upr      p adj
## Relaxed-Friendly      0.4309000 -0.4200711 1.281871 0.4435799
## Temperamental-Friendly 1.0056429  0.2453111 1.765975 0.0068745
## Temperamental-Relaxed  0.5747429 -0.3613579 1.510844 0.3061736
```

```
# Other post-hoc tests use the pairwise.t.test function; but
# you need to use source$variable notation instead of
# tidyverse style.
pairwise.t.test(
  df$volunteering,  # Outcome variable
  df$psychRegions,  # Grouping variable
  p.adj = "bonf"    # Adjustment method
)
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  df$volunteering and df$psychRegions
##
##              Friendly Relaxed
## Relaxed      0.6784   -
## Temperamental 0.0074   0.4311
##
## P value adjustment method: bonferroni
```

```
# Information on methods available for post-hoc tests
?p.adjust
```