Visualizations

Diana imali Luvandale

2023-08-30

Introduction

We have been given a dataset which contains which contains data collected via a questionnaire from Kenyan high schools in June 2018. The primary objective is to determine if a relationship exists between Generalized Anxiety Disorders and demographics such as tribe, age, gender and school.

The dataset contains the following variables:

- Patient Health Questionnaire-8 (PHQ-8)
- Generalized Anxiety Disorder Screener-7 (GAD-7)
- Multidimensional Scale of Perceived Social Support (MSSS)
- Tribe
- Gender
- Age
- School
- School Resources

We would like to first explore the data

Data Loading and Exploration

The data was obtained from the link. The data initially contained 659 observations (rows) and 33 variables (columns).

```
#loading the necessary libraries

library(epitools)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
dat1 <-read.csv("/home/imalizzz/Downloads/shamiri_imputed_dataset.csv")
head(dat1)
```

```
##   ParticipantID PHQ1 PHQ2 PHQ3 PHQ4 PHQ5 PHQ6 PHQ7 PHQ8 GAD1 GAD2 GAD3 GAD4
## 1        SR_001    3    0    2    1    1    1    3    1    1    2    3    3
## 2        SR_002    3    0    1    0    0    1    2    2    0    0    0    1
## 3        SR_003    2    3    0    1    2    3    1    2    0    2    3    1
## 4        SR_004    1    3    1    1    2    1    3    1    3    3    3    1
## 5        SR_005    1    1    0    1    0    3    3    0    1    3    3    1
## 6        SR_006    0    2    0    2    0    0    2    0    2    0    2    0
##   GAD5 GAD6 GAD7 MSSS1 MSSS2 MSSS3 MSSS4 MSSS5 MSSS6 MSSS7 MSSS8 MSSS9 MSSS10
## 1    1    3    0     1     1     6     5     1     5     6     7     6      7
## 2    1    0    0     1     4     6     5     5     3     2     7     5      7
## 3    0    3    1     2     2     7     5     2     5     2     5     2      2
## 4    1    3    2     4     4     5     4     5     4     3     4     4      5
## 5    0    1    0     6     6     6     6     6     2     2     6     6      2
## 6    2    0    2     7     6     7     6     7     2     4     7     6      5
##   MSSS11 MSSS12    Tribe Gender  School Age School_Resources
## 1      6      6 Minority      M Starays  18             Poor
## 2      3      6 Minority      M Starays  16             Poor
## 3      5      5 Minority      F Starays  14             Poor
## 4      4      3 Minority      M Starays  20             Poor
## 5      6      2 Minority      M Starays  18             Poor
## 6      6      7 Minority      F Starays  16             Poor
```

Data Preprocessing

checking the summary of the dataset.This shows the summary of every column including the mean of every GAD

```r
#checking the attributes of the dataset
attributes(dat1)
```

```
## $names
##  [1] "ParticipantID"    "PHQ1"             "PHQ2"             "PHQ3"
##  [5] "PHQ4"             "PHQ5"             "PHQ6"             "PHQ7"
##  [9] "PHQ8"             "GAD1"             "GAD2"             "GAD3"
## [13] "GAD4"             "GAD5"             "GAD6"             "GAD7"
## [17] "MSSS1"            "MSSS2"            "MSSS3"            "MSSS4"
## [21] "MSSS5"            "MSSS6"            "MSSS7"            "MSSS8"
## [25] "MSSS9"            "MSSS10"           "MSSS11"           "MSSS12"
## [29] "Tribe"            "Gender"           "School"           "Age"
## [33] "School_Resources"
##
## $class
## [1] "data.frame"
##
## $row.names
##   [1]   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##  [19]  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##  [37]  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##  [55]  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##  [73]  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##  [91]  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108
## [109] 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
## [127] 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
## [145] 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162
## [163] 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
## [181] 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198
## [199] 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
## [217] 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234
## [235] 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252
## [253] 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
## [271] 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288
## [289] 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306
## [307] 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324
## [325] 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342
## [343] 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
## [361] 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378
## [379] 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396
## [397] 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
## [415] 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432
## [433] 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
## [451] 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468
## [469] 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486
## [487] 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504
## [505] 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522
## [523] 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
## [541] 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558
## [559] 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576
## [577] 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594
## [595] 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612
## [613] 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
```

```
## [631] 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648
## [649] 649 650 651 652 653 654 655 656 657 658
```

```
summary(dat1)
```

```
##  ParticipantID           PHQ1            PHQ2            PHQ3
##  Length:658         Min.   :0.00   Min.   :0.000   Min.   :0.000
##  Class :character   1st Qu.:0.00   1st Qu.:0.000   1st Qu.:0.000
##  Mode  :character   Median :1.00   Median :1.000   Median :1.000
##                     Mean   :1.33   Mean   :1.217   Mean   :1.153
##                     3rd Qu.:2.00   3rd Qu.:2.000   3rd Qu.:2.000
##                     Max.   :3.00   Max.   :3.000   Max.   :4.000
##       PHQ4            PHQ5             PHQ6            PHQ7
##  Min.   :0.000   Min.   :0.0000   Min.   :0.000   Min.   :0.00
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:1.00
##  Median :1.000   Median :1.0000   Median :1.000   Median :1.00
##  Mean   :1.149   Mean   :0.9331   Mean   :1.277   Mean   :1.41
##  3rd Qu.:2.000   3rd Qu.:2.0000   3rd Qu.:2.000   3rd Qu.:3.00
##  Max.   :3.000   Max.   :3.0000   Max.   :3.000   Max.   :3.00
##       PHQ8             GAD1            GAD2            GAD3
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:1.000
##  Median :0.0000   Median :1.000   Median :1.000   Median :1.000
##  Mean   :0.7568   Mean   :1.049   Mean   :1.319   Mean   :1.515
##  3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:3.000
##  Max.   :3.0000   Max.   :3.000   Max.   :3.000   Max.   :3.000
##       GAD4             GAD5             GAD6            GAD7
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000
##  Median :1.0000   Median :0.0000   Median :1.000   Median :1.000
##  Mean   :0.9985   Mean   :0.7766   Mean   :1.204   Mean   :1.257
##  3rd Qu.:2.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :3.0000   Max.   :3.0000   Max.   :3.000   Max.   :3.000
##      MSSS1           MSSS2           MSSS3           MSSS4           MSSS5
##  Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.00   1st Qu.:4.000   1st Qu.:6.000   1st Qu.:5.000   1st Qu.:4.000
##  Median :6.00   Median :6.000   Median :6.000   Median :6.000   Median :6.000
##  Mean   :5.12   Mean   :5.167   Mean   :5.843   Mean   :5.305   Mean   :5.289
##  3rd Qu.:7.00   3rd Qu.:7.000   3rd Qu.:7.000   3rd Qu.:7.000   3rd Qu.:7.000
##  Max.   :7.00   Max.   :7.000   Max.   :7.000   Max.   :7.000   Max.   :7.000
##      MSSS6           MSSS7           MSSS8           MSSS9           MSSS10
##  Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.000   1st Qu.:2.000   1st Qu.:4.00   1st Qu.:4.000   1st Qu.:4.000
##  Median :5.000   Median :5.000   Median :5.00   Median :5.000   Median :6.000
##  Mean   :4.556   Mean   :4.271   Mean   :4.88   Mean   :4.854   Mean   :5.153
##  3rd Qu.:6.000   3rd Qu.:6.000   3rd Qu.:6.00   3rd Qu.:6.000   3rd Qu.:7.000
##  Max.   :7.000   Max.   :7.000   Max.   :7.00   Max.   :7.000   Max.   :8.000
##      MSSS11          MSSS12          Tribe              Gender
##  Min.   :1.00   Min.   :1.000   Length:658         Length:658
##  1st Qu.:5.00   1st Qu.:3.000   Class :character   Class :character
##  Median :6.00   Median :5.000   Mode  :character   Mode  :character
##  Mean   :5.45   Mean   :4.207
##  3rd Qu.:7.00   3rd Qu.:6.000
##  Max.   :7.00   Max.   :7.000
```

```
##     School               Age        School_Resources
##  Length:658         Min.   :12.00   Length:658
##  Class :character   1st Qu.:15.00   Class :character
##  Mode  :character   Median :16.00   Mode  :character
##                     Mean   :15.85
##                     3rd Qu.:17.00
##                     Max.   :25.00
```

Checking for misting values. The data has no missing values as seen below

```
#checking the no of missing values in each column
colSums(is.na(dat1)) # This shows that there are no missing values in any of the columns
```

```
##    ParticipantID             PHQ1             PHQ2             PHQ3
##                0                0                0                0
##             PHQ4             PHQ5             PHQ6             PHQ7
##                0                0                0                0
##             PHQ8             GAD1             GAD2             GAD3
##                0                0                0                0
##             GAD4             GAD5             GAD6             GAD7
##                0                0                0                0
##            MSSS1            MSSS2            MSSS3            MSSS4
##                0                0                0                0
##            MSSS5            MSSS6            MSSS7            MSSS8
##                0                0                0                0
##            MSSS9           MSSS10           MSSS11           MSSS12
##                0                0                0                0
##            Tribe           Gender           School              Age
##                0                0                0                0
## School_Resources
##                0
```

Finding the unique values in each column. Doing this to check if respondents included only the provided choice ie 0-3. Other values could affect statistical computations

```
rapply(dat1,function(x)length(unique(x)))
```

```
##    ParticipantID             PHQ1             PHQ2             PHQ3
##              626                4                4                5
##             PHQ4             PHQ5             PHQ6             PHQ7
##                4                4                4                4
##             PHQ8             GAD1             GAD2             GAD3
##                4                4                4                4
##             GAD4             GAD5             GAD6             GAD7
##                4                4                4                4
##            MSSS1            MSSS2            MSSS3            MSSS4
##                7                7                7                7
##            MSSS5            MSSS6            MSSS7            MSSS8
##                7                7                7                7
##            MSSS9           MSSS10           MSSS11           MSSS12
##                7                8                7                7
##            Tribe           Gender           School              Age
```

```
##                2                2                5               14
## School_Resources
##                3
```

Realized PHQ3 has option 4 which could be a data entry problem. The 4 will be dropped to assist with statistics computations later on

```
categories <- unique(dat1$PHQ3)
categories
```

```
## [1] 2 1 0 3 4
```

EDA(Exploratory Data analysis)

Overall mean scores on the GAD

```
#finding the overall meanscores of the GAD
overall_mean <- colMeans(dat1[ , c(10, 11,12,13,14,15,16)])
overall_mean
```

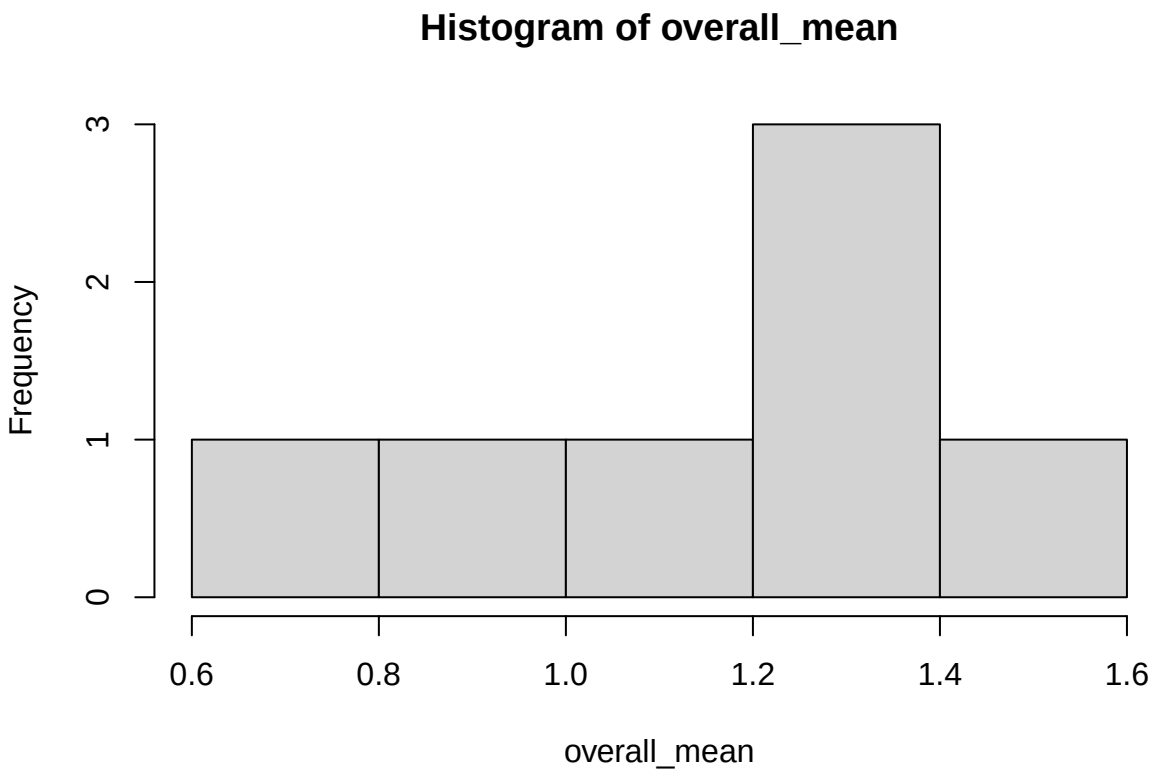
```
##      GAD1      GAD2      GAD3      GAD4      GAD5      GAD6      GAD7
## 1.0486322 1.3191489 1.5151976 0.9984802 0.7765957 1.2036474 1.2568389
```

Histogram visualization of the overall GAD mean

```
#converting to data frame
data=data.frame(overall_mean)

print(data)
```

```
##      overall_mean
## GAD1    1.0486322
## GAD2    1.3191489
## GAD3    1.5151976
## GAD4    0.9984802
## GAD5    0.7765957
## GAD6    1.2036474
## GAD7    1.2568389
```

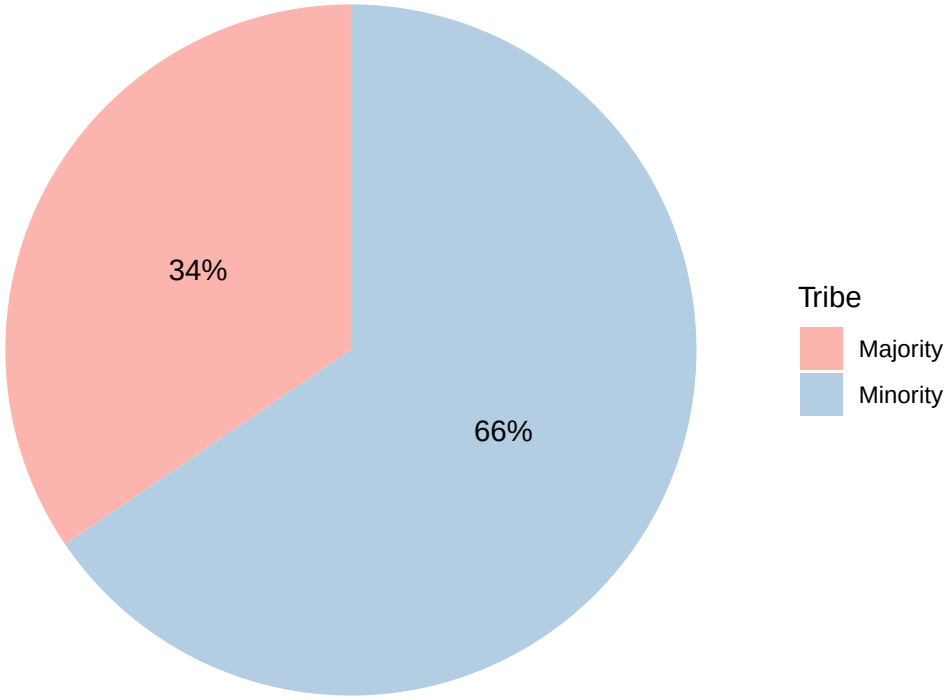
```
#histogram of overall mean
hist(overall_mean)
```



Different visualizations including gender, tribe, age, school and school resources distribution

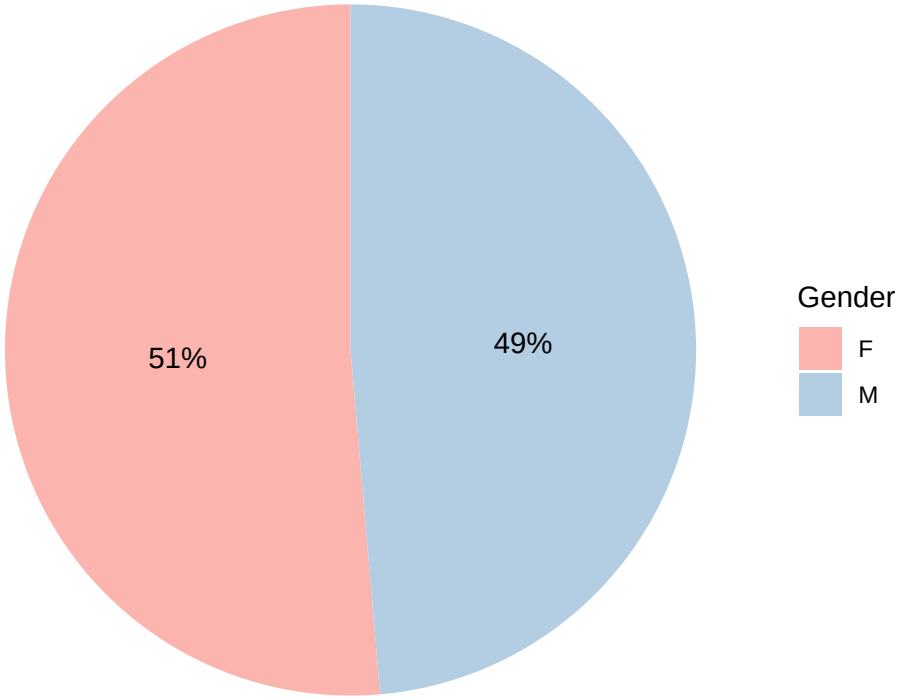
Majority make up 66% of the respondents while minority make up 34%

```
#Tribe visualization
ggplot(dat1, aes("",fill = Tribe)) +
  geom_bar(position="fill") +
  geom_text(
    stat='count',
    aes(y=after_stat(..count..),
        label=after_stat(scales::percent(..count../sum(..count..),1))),
    position=position_fill(0.5),
  ) +
  coord_polar(theta="y") +
  labs(x=NULL, y=NULL) +
  scale_fill_brewer(palette="Pastel1") +
  theme_void()
```



The respondents were 51% female and 49% male

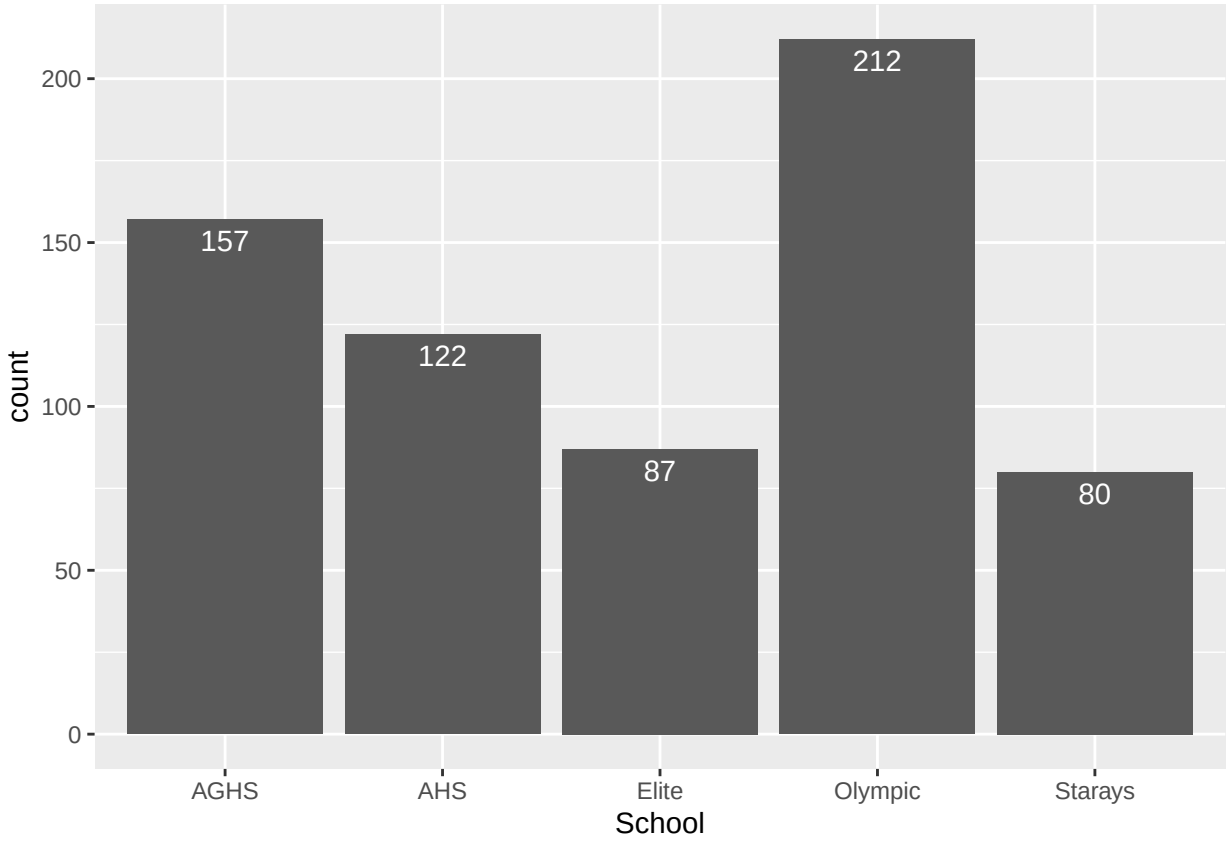
```
#Gender visualization

ggplot(dat1, aes("",fill = Gender)) +
  geom_bar(position="fill") +
  geom_text(
    stat='count',
    aes(y=after_stat(..count..),
        label=after_stat(scales::percent(..count../sum(..count..),1))),
    position=position_fill(0.5),
  ) +
  coord_polar(theta="y") +
  labs(x=NULL, y=NULL) +
  scale_fill_brewer(palette="Pastel1") +
  theme_void()
```



The school with the highest no of respondents was Olympic, followed by AGHS

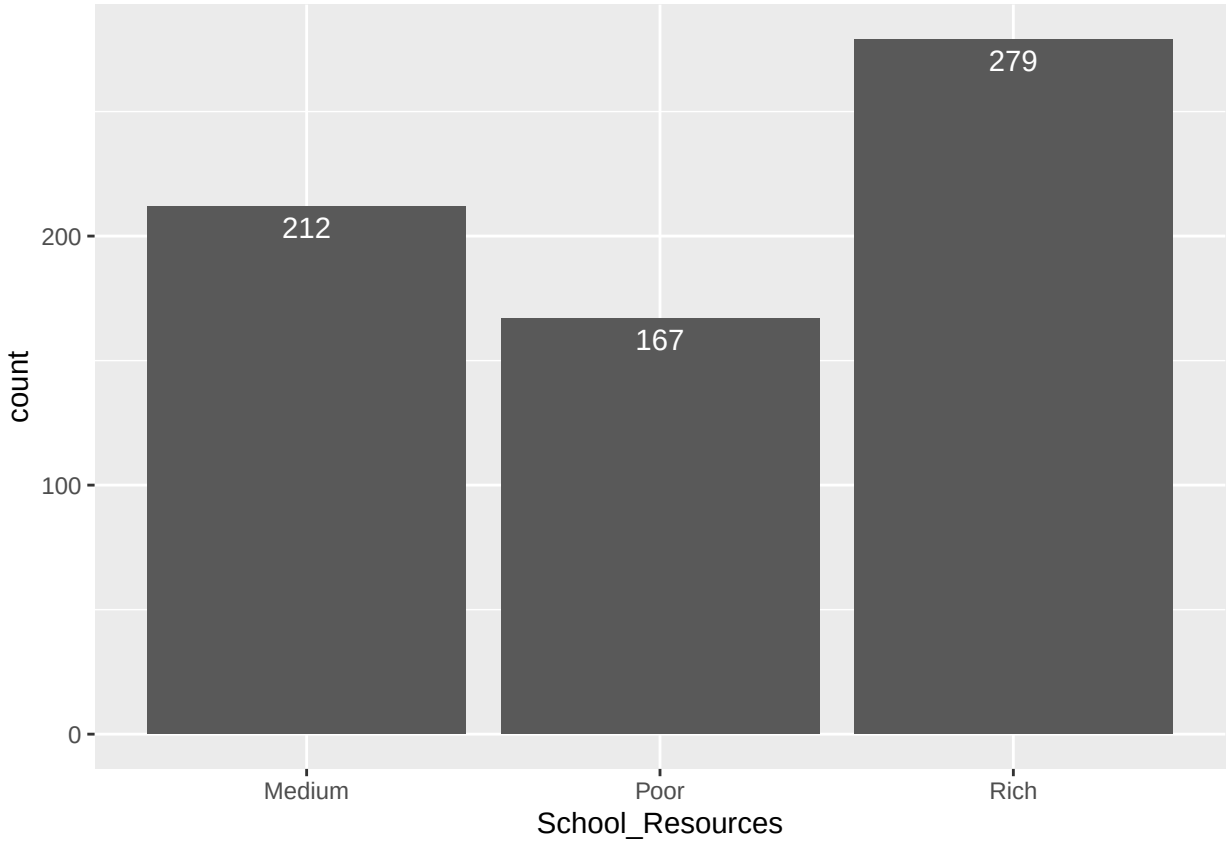
```
#School visualization

ggplot(dat1, aes(x = School)) +
  geom_bar() +
  geom_text(aes(label = ..count..), stat = "count", vjust = 1.5, colour = "white")
```

```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(count)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



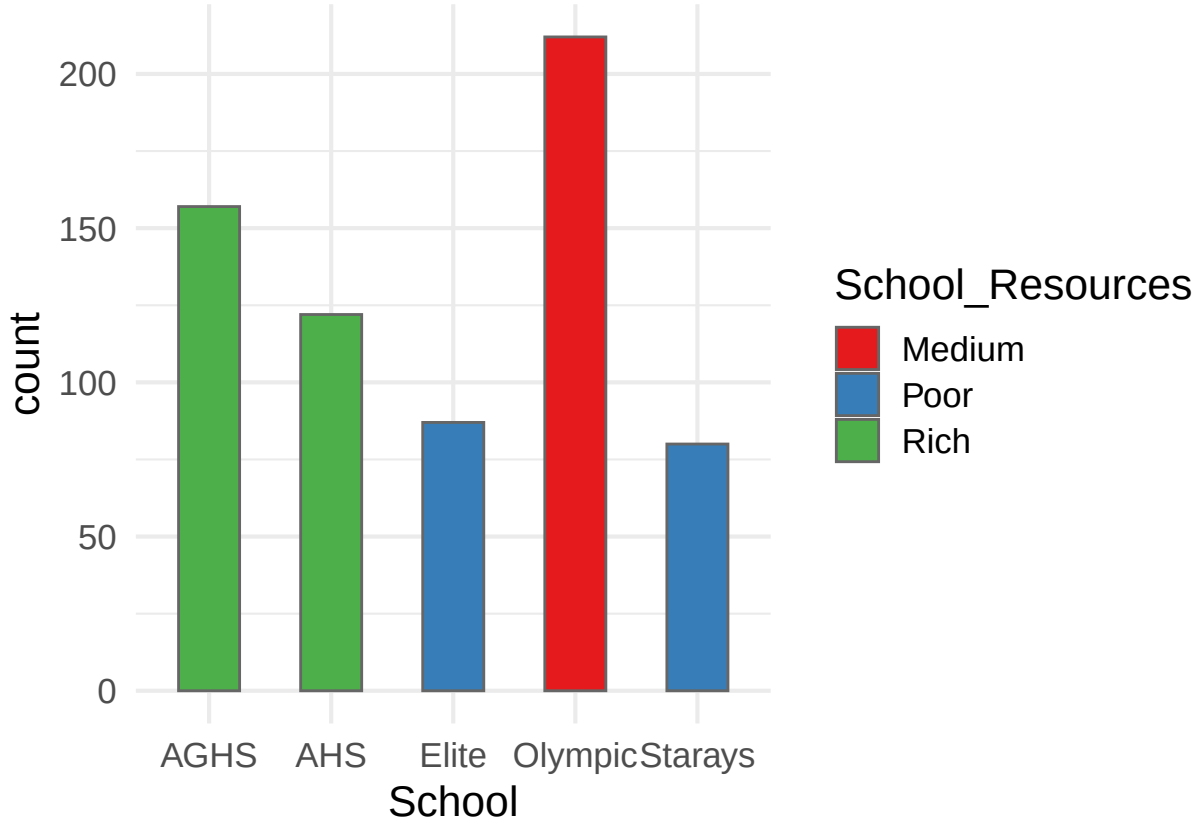
There are more rich school resources

```
#School resources

ggplot(dat1, aes(x = School_Resources)) +
  geom_bar() +
  geom_text(aes(label = ..count..), stat = "count", vjust = 1.5, colour = "white")
```



Later on we realize schools with poor resources exhibited higher levels of anxiety

```
#School resources vs School
ggplot(dat1, aes(x = School, fill = School_Resources)) +
  geom_bar(width = 0.5, color = 'gray40') +
  scale_fill_brewer(palette = 'Set1') +
  theme_minimal(base_size = 16)
```



GAD ANALYSIS

Tribe GAD Mean

Minority had the highest mean in GAD1,GAD2,GAD3,GAD5,GAD6,GAD7, Except for GAD4

```
df2 <- dat1 %>% group_by(Tribe) %>%
  summarise(mean_GAD1=mean(GAD1),
            mean_GAD2= mean(GAD2),mean_GAD3=mean(GAD3),mean_GAD4=mean(GAD4),mean_GAD5=mean(GAD5),mean_GA
            .groups = 'drop') %>%
  as.data.frame()
df2
```

```
##      Tribe mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6
## 1 Majority 0.9295154  1.176211  1.325991 1.0220264 0.6475771  1.145374
## 2 Minority 1.1113689  1.394432  1.614849 0.9860789 0.8445476  1.234339
##   mean_GAD7
## 1  1.162996
## 2  1.306265
```

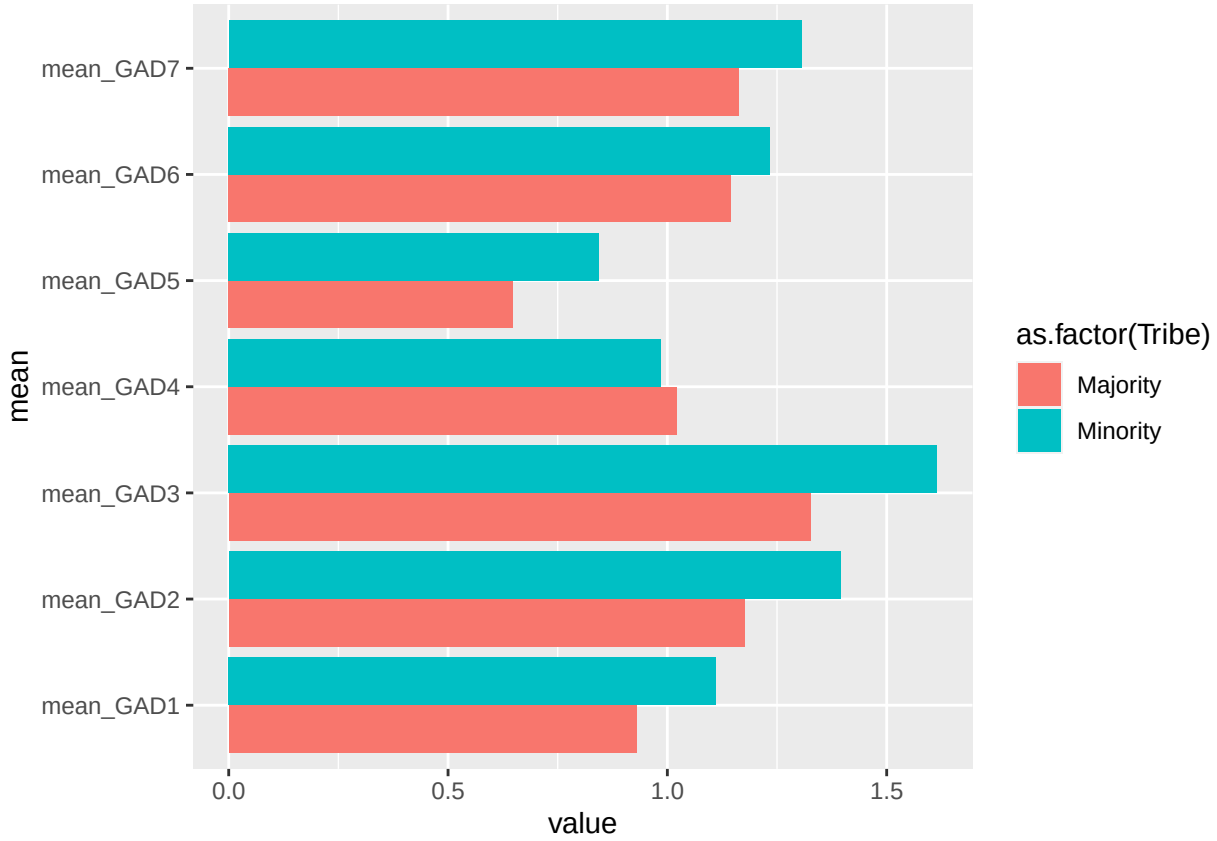
```
# Load the package
library(tidyr)
library(dplyr)
```

```r
library(knitr)

# transform the format of the data to help in easier visualization
data_long <- gather(df2, mean, value, mean_GAD1:mean_GAD7) %>%
  arrange(factor(Tribe, levels = c("Majority", "Minority"))) %>%
  mutate(Tribe=factor(Tribe, levels=unique(Tribe)))


kable(head(data_long, 10))
```

Tribe	mean	value
Majority	mean_GAD1	0.9295154
Majority	mean_GAD2	1.1762115
Majority	mean_GAD3	1.3259912
Majority	mean_GAD4	1.0220264
Majority	mean_GAD5	0.6475771
Majority	mean_GAD6	1.1453744
Majority	mean_GAD7	1.1629956
Minority	mean_GAD1	1.1113689
Minority	mean_GAD2	1.3944316
Minority	mean_GAD3	1.6148492

```r
data_long %>%
  ggplot(aes(x = value, y = mean, fill = as.factor(Tribe))) +
  geom_bar(stat = "identity", position = "dodge",space = 2 )
```

```
## Warning in geom_bar(stat = "identity", position = "dodge", space = 2): Ignoring
## unknown parameters: ‘space‘
```



Finding the overall mean of tribe. Majority has an overall GAD mean of 1.058527 and minority 1.213126

```
df2_r <- subset(df2, select = -Tribe)
df2_r
```

```
##   mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1 0.9295154  1.176211  1.325991 1.0220264 0.6475771  1.145374  1.162996
## 2 1.1113689  1.394432  1.614849 0.9860789 0.8445476  1.234339  1.306265
```

```
rowMeans(df2_r)
```

```
##        1        2
## 1.058527 1.213126
```

gender

females have the highest means of GAD from 1 to 7 compared to their male counterparts

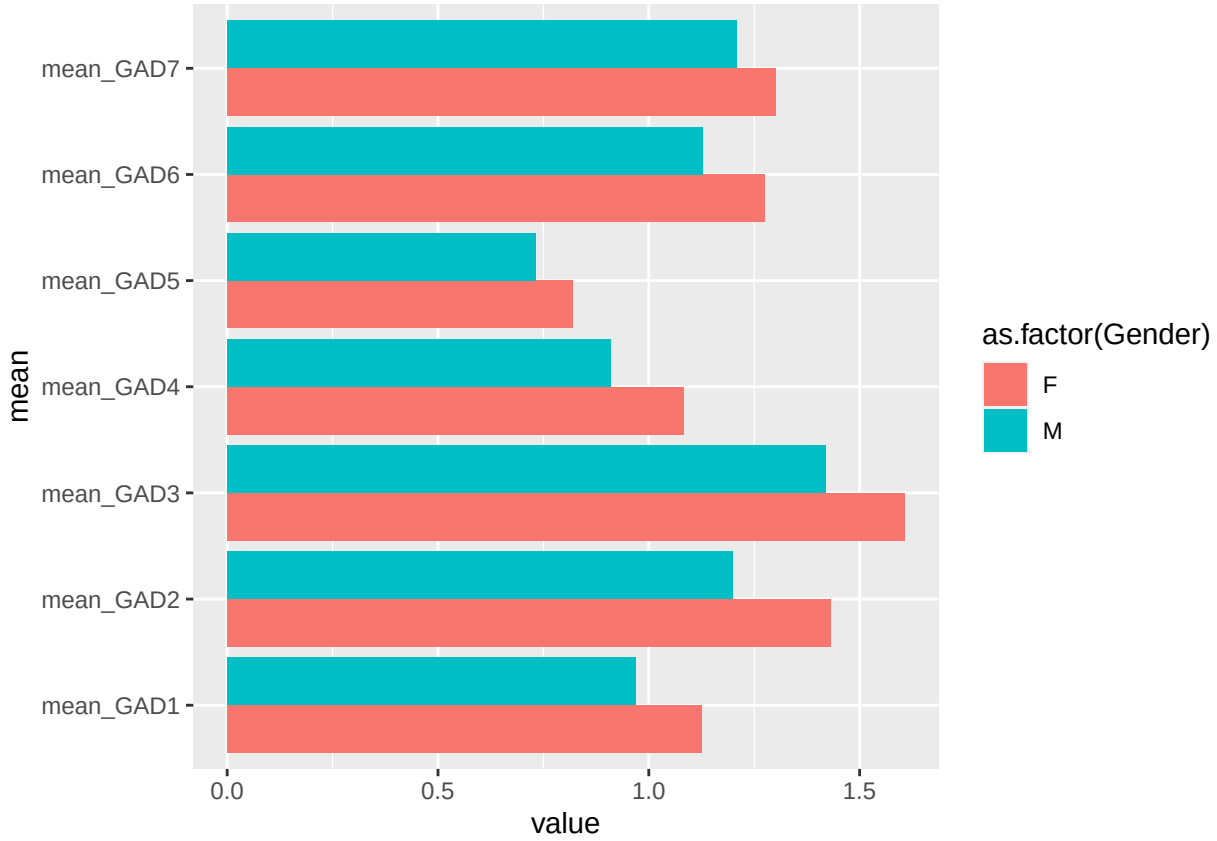
```
df3 <- dat1 %>% group_by(Gender) %>%
  summarise(mean_GAD1=mean(GAD1),
            mean_GAD2= mean(GAD2),mean_GAD3=mean(GAD3),mean_GAD4=mean(GAD4),mean_GAD5=mean(GAD5),mean_G
            .groups = 'drop') %>%
  as.data.frame()
df3
```

```
##   Gender mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1      F   1.12426  1.431953  1.606509  1.082840 0.8195266  1.275148  1.301775
## 2      M   0.96875  1.200000  1.418750  0.909375 0.7312500  1.128125  1.209375
```

```r
# transform the format
df3_long <- gather(df3, mean, value, mean_GAD1:mean_GAD7) %>%
  arrange(factor(Gender, levels = c("Female", "Male"))) %>%
  mutate(Gender=factor(Gender, levels=unique(Gender)))

kable(head(df3_long, 10))
```

Gender	mean	value
F	mean_GAD1	1.1242604
M	mean_GAD1	0.9687500
F	mean_GAD2	1.4319527
M	mean_GAD2	1.2000000
F	mean_GAD3	1.6065089
M	mean_GAD3	1.4187500
F	mean_GAD4	1.0828402
M	mean_GAD4	0.9093750
F	mean_GAD5	0.8195266
M	mean_GAD5	0.7312500

```r
#Bar graph mean visualizations
df3_long %>%
  ggplot(aes(x = value, y = mean, fill = as.factor(Gender))) +
  geom_bar(stat = "identity", position = "dodge")
```



Finding the overall mean of Gender . Female has a mean of 1.234573 and male 1.080804

```
df3_r <- subset(df3, select = -Gender)
df3_r
```

```
##   mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1   1.12426  1.431953  1.606509  1.082840 0.8195266  1.275148  1.301775
## 2   0.96875  1.200000  1.418750  0.909375 0.7312500  1.128125  1.209375
```

```
rowMeans(df3_r)
```

```
##        1        2
## 1.234573 1.080804
```

School

Starays has the highest GAD1,GAD4; AGHS has the highest GAD2; Elite has the highest GAD3, GAD5,GAD6,GAD7

```
#school
df4 <- dat1 %>% group_by(School) %>%
  summarise(mean_GAD1=mean(GAD1),
            mean_GAD2= mean(GAD2),mean_GAD3=mean(GAD3),mean_GAD4=mean(GAD4),mean_GAD5=mean(GAD5),mean_GA
```

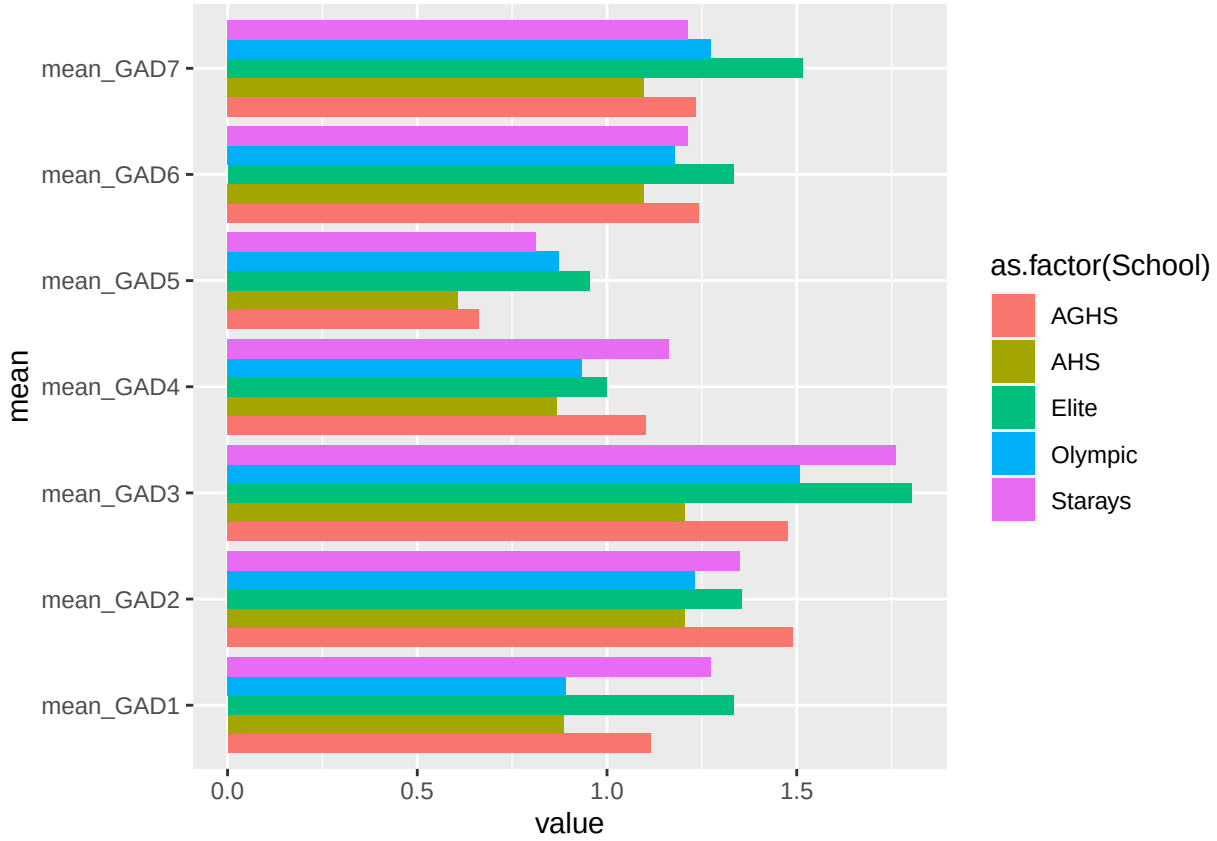
```
            .groups = 'drop') %>%
  as.data.frame()
df4
```

```
##    School mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1    AGHS 1.1146497  1.490446  1.477707 1.1019108 0.6624204  1.242038  1.235669
## 2     AHS 0.8852459  1.204918  1.204918 0.8688525 0.6065574  1.098361  1.098361
## 3   Elite 1.3333333  1.356322  1.804598 1.0000000 0.9540230  1.333333  1.517241
## 4 Olympic 0.8915094  1.231132  1.509434 0.9339623 0.8726415  1.179245  1.273585
## 5 Starays 1.2750000  1.350000  1.762500 1.1625000 0.8125000  1.212500  1.212500
```

```
# transform the format
df4_long <- gather(df4, mean, value, mean_GAD1:mean_GAD7) %>%
  arrange(factor(School, levels = c("AGHS", "AHS", "Elite","Olympic","Starays"))) %>%
  mutate(School=factor(School, levels=unique(School)))

kable(head(df4_long, 10))
```

School	mean	value
AGHS	mean_GAD1	1.1146497
AGHS	mean_GAD2	1.4904459
AGHS	mean_GAD3	1.4777070
AGHS	mean_GAD4	1.1019108
AGHS	mean_GAD5	0.6624204
AGHS	mean_GAD6	1.2420382
AGHS	mean_GAD7	1.2356688
AHS	mean_GAD1	0.8852459
AHS	mean_GAD2	1.2049180
AHS	mean_GAD3	1.2049180

```
df4_long %>%
  ggplot(aes(x = value, y = mean, fill = as.factor(School))) +
  geom_bar(stat = "identity", position = "dodge")
```



Overall GAD mean of schools. Elite has the highest overall GAD mean of 1.3284072, followed by Olympic 1.1273585. AHS has the lowest 0.9953162

```
df4_r <- subset(df4, select = -School)
df4_r
```

```
##   mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1 1.1146497  1.490446  1.477707 1.1019108 0.6624204  1.242038  1.235669
## 2 0.8852459  1.204918  1.204918 0.8688525 0.6065574  1.098361  1.098361
## 3 1.3333333  1.356322  1.804598 1.0000000 0.9540230  1.333333  1.517241
## 4 0.8915094  1.231132  1.509434 0.9339623 0.8726415  1.179245  1.273585
## 5 1.2750000  1.350000  1.762500 1.1625000 0.8125000  1.212500  1.212500
```

```
rowMeans(df4_r)
```

```
##         1         2         3         4         5
## 1.1892630 0.9953162 1.3284072 1.1273585 1.2553571
```

Age

22 and 25 yr old have the highest GAD score

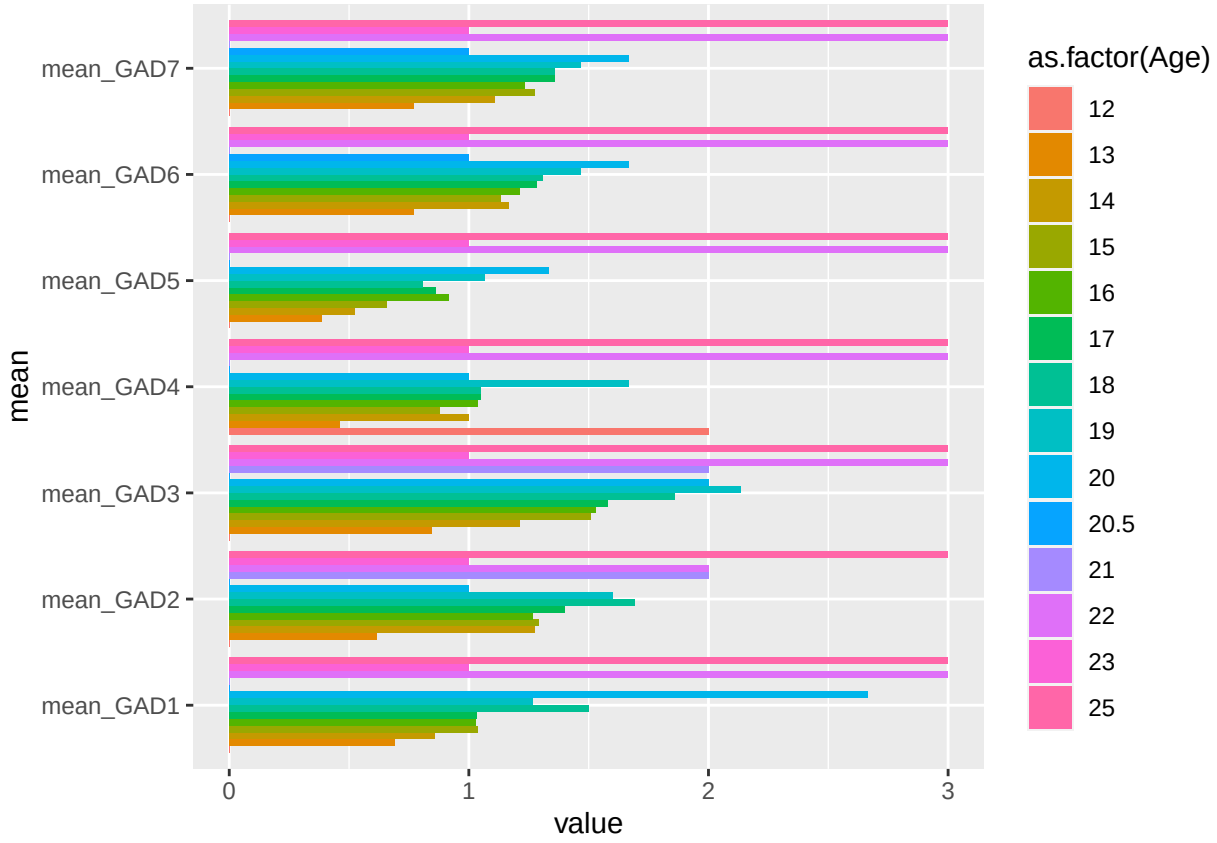
```r
df5 <- dat1 %>% group_by(Age) %>%
  summarise(mean_GAD1=mean(GAD1),
            mean_GAD2= mean(GAD2),mean_GAD3=mean(GAD3),mean_GAD4=mean(GAD4),mean_GAD5=mean(GAD5),mean_GA
            .groups = 'drop') %>%
  as.data.frame()
df5
```

```
##     Age mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1  12.0 0.0000000 0.0000000 0.0000000 2.0000000 0.0000000 0.0000000 0.0000000
## 2  13.0 0.6923077 0.6153846 0.8461538 0.4615385 0.3846154 0.7692308 0.7692308
## 3  14.0 0.8571429 1.2738095 1.2142857 1.0000000 0.5238095 1.1666667 1.1071429
## 4  15.0 1.0391061 1.2905028 1.5083799 0.8770950 0.6592179 1.1340782 1.2737430
## 5  16.0 1.0301508 1.2663317 1.5276382 1.0351759 0.9145729 1.2110553 1.2311558
## 6  17.0 1.0341880 1.4017094 1.5811966 1.0512821 0.8632479 1.2820513 1.3589744
## 7  18.0 1.5000000 1.6904762 1.8571429 1.0476190 0.8095238 1.3095238 1.3571429
## 8  19.0 1.2666667 1.6000000 2.1333333 1.6666667 1.0666667 1.4666667 1.4666667
## 9  20.0 2.6666667 1.0000000 2.0000000 1.0000000 1.3333333 1.6666667 1.6666667
## 10 20.5 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 1.0000000 1.0000000
## 11 21.0 0.0000000 2.0000000 2.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## 12 22.0 3.0000000 2.0000000 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000
## 13 23.0 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## 14 25.0 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000
```

```r
# transform the format
df5_long <- gather(df5, mean, value, mean_GAD1:mean_GAD7) %>%
  arrange(factor(Age, levels = c(12,13,14,15,16,17,18,19,20,20.5,21,22,23,25))) %>%
  mutate(Age=factor(Age, levels=unique(Age)))

kable(head(df5_long, 10))
```

Age	mean	value
12	mean_GAD1	0.0000000
12	mean_GAD2	0.0000000
12	mean_GAD3	0.0000000
12	mean_GAD4	2.0000000
12	mean_GAD5	0.0000000
12	mean_GAD6	0.0000000
12	mean_GAD7	0.0000000
13	mean_GAD1	0.6923077
13	mean_GAD2	0.6153846
13	mean_GAD3	0.8461538

```r
df5_long %>%
  ggplot(aes(x = value, y = mean, fill = as.factor(Age))) +
  geom_bar(stat = "identity", position = "dodge")
```



Overall mean GAD for ages. Ages 22 and 25 have the highest GAD means for all except GAD 2

```
df5_r <- subset(df5, select = -Age)
df5_r
```

```
##    mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6 mean_GAD7
## 1  0.0000000 0.0000000 0.0000000 2.0000000 0.0000000 0.0000000 0.0000000
## 2  0.6923077 0.6153846 0.8461538 0.4615385 0.3846154 0.7692308 0.7692308
## 3  0.8571429 1.2738095 1.2142857 1.0000000 0.5238095 1.1666667 1.1071429
## 4  1.0391061 1.2905028 1.5083799 0.8770950 0.6592179 1.1340782 1.2737430
## 5  1.0301508 1.2663317 1.5276382 1.0351759 0.9145729 1.2110553 1.2311558
## 6  1.0341880 1.4017094 1.5811966 1.0512821 0.8632479 1.2820513 1.3589744
## 7  1.5000000 1.6904762 1.8571429 1.0476190 0.8095238 1.3095238 1.3571429
## 8  1.2666667 1.6000000 2.1333333 1.6666667 1.0666667 1.4666667 1.4666667
## 9  2.6666667 1.0000000 2.0000000 1.0000000 1.3333333 1.6666667 1.6666667
## 10 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 1.0000000 1.0000000
## 11 0.0000000 2.0000000 2.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## 12 3.0000000 2.0000000 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000
## 13 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## 14 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000 3.0000000
```

```
rowMeans(df5_r)
```

```
##         1         2         3         4         5         6         7         8
## 0.2857143 0.6483516 1.0204082 1.1117318 1.1737258 1.2246642 1.3673469 1.5238095
```

```
##         9        10        11        12        13        14
## 1.6190476 0.2857143 0.5714286 2.8571429 1.0000000 3.0000000
```

Resources

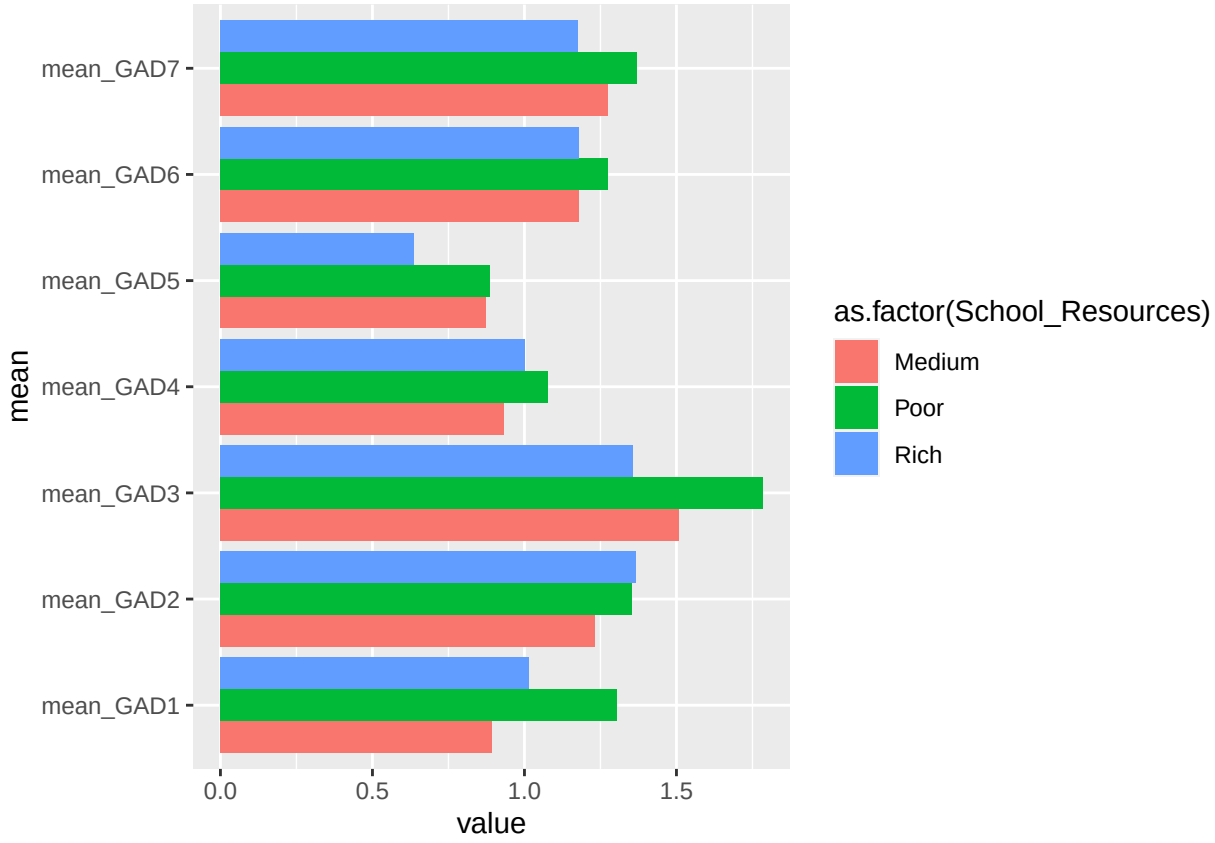
```
df6 <- dat1 %>% group_by( School_Resources) %>%
  summarise(mean_GAD1=mean(GAD1),
            mean_GAD2= mean(GAD2),mean_GAD3=mean(GAD3),mean_GAD4=mean(GAD4),mean_GAD5=mean(GAD5),mean_GA
            .groups = 'drop') %>%
  as.data.frame()
df6
```

```
##   School_Resources mean_GAD1 mean_GAD2 mean_GAD3 mean_GAD4 mean_GAD5 mean_GAD6
## 1           Medium 0.8915094  1.231132  1.509434 0.9339623 0.8726415  1.179245
## 2             Poor 1.3053892  1.353293  1.784431 1.0778443 0.8862275  1.275449
## 3             Rich 1.0143369  1.365591  1.358423 1.0000000 0.6379928  1.179211
##   mean_GAD7
## 1  1.273585
## 2  1.371257
## 3  1.175627
```

```
# transform the format
df6_long <- gather(df6, mean, value, mean_GAD1:mean_GAD7) %>%
  arrange(factor(School_Resources, levels = c("Medium", "Poor", "Rich"))) %>%
  mutate(School_Resources=factor(School_Resources, levels=unique(School_Resources)))

kable(head(df6_long, 10))
```

School_Resources	mean	value
Medium	mean_GAD1	0.8915094
Medium	mean_GAD2	1.2311321
Medium	mean_GAD3	1.5094340
Medium	mean_GAD4	0.9339623
Medium	mean_GAD5	0.8726415
Medium	mean_GAD6	1.1792453
Medium	mean_GAD7	1.2735849
Poor	mean_GAD1	1.3053892
Poor	mean_GAD2	1.3532934
Poor	mean_GAD3	1.7844311

```
df6_long %>%
  ggplot(aes(x = value, y = mean, fill = as.factor(School_Resources))) +
  geom_bar(stat = "identity", position = "dodge")
```


mean_GAD7
mean_GAD6
mean_GAD5
mean_GAD4
mean_GAD3
mean_GAD2
mean_GAD1
mean
0.0
0.5
1.0
1.5
value
as.factor(School_Resources)
Medium
Poor
Rich