

# Transformations

Ryan T. Moore

2025-01-04

## Prelim

```
library(dplyr)
library(forcats)
data("gss-cat")
```

### 1. Sort by tvhours

```
gss_cat <- gss_cat |> arrange(desc(tvhours))
```

### 2. Create birthyear, show summary and dims

```
gss_cat <- gss_cat |> mutate(birthyear = year - age)

# Summarise new df:
summary(gss_cat)
```

```
##      year      marital      age      race
## Min.   :2000   No answer   : 17   Min.   :18.00   Other      : 1959
## 1st Qu.:2002   Never married: 5416   1st Qu.:33.00   Black      : 3129
## Median :2006   Separated   : 743   Median :46.00   White      :16395
## Mean   :2007   Divorced    : 3383   Mean   :47.18   Not applicable: 0
## 3rd Qu.:2010   Widowed     : 1807   3rd Qu.:59.00
## Max.   :2014   Married     :10117   Max.   :89.00
##                      NA's    :76
##      rincome      partyid      relig
## $25000 or more:7363   Independent   :4119   Protestant:10846
## Not applicable:7043   Not str democrat :3690   Catholic  : 5124
## $20000 - 24999:1283   Strong democrat :3490   None      : 3523
## $10000 - 14999:1168   Not str republican:3032   Christian : 689
## $15000 - 19999:1048   Ind,near dem    :2499   Jewish    : 388
## Refused       : 975   Strong republican :2314   Other     : 224
## (Other)       :2603   (Other)         :2339   (Other)   : 689
```

```
##           denom      tvhours      birthyear
## Not applicable :10072   Min.    : 0.000   Min.    :1911
## Other          : 2534   1st Qu.: 1.000   1st Qu.:1947
## No denomination : 1683   Median : 2.000   Median :1960
## Southern baptist: 1536   Mean    : 2.981   Mean    :1959
## Baptist-dk which: 1457   3rd Qu.: 4.000   3rd Qu.:1973
## United methodist: 1067   Max.    :24.000   Max.    :1996
## (Other)        : 3134   NA's    :10146   NA's    :76
```

```
dim(gss_cat)
```

```
## [1] 21483    10
```

Now, `gss_cat` has 21483 rows.

### 3. Only tvhours > 3

```
gss_cat_tv <- gss_cat |> filter(tvhours > 3)
gss_cat_tv
```

```
## # A tibble: 3,318 x 10
##   year marital      age race  rincome partyid relig denom tvhours birthyear
##   <int> <fct>      <int> <fct> <fct>   <fct>   <fct> <fct>   <int>    <int>
## 1  2000 Never married  30 Black Not ap~ Indepe~ Prot~ Bapt~    24     1970
## 2  2000 Separated     45 Black Not ap~ Ind,ne~ Prot~ Other    24     1955
## 3  2002 Never married  33 White $6000 ~ Indepe~ Cath~ Not ~    24     1969
## 4  2006 Divorced     53 Black Not ap~ Strong~ Prot~ Sout~    24     1953
## 5  2008 Divorced     50 Black No ans~ Ind,ne~ Prot~ Bapt~    24     1958
## 6  2008 Never married  44 White Not ap~ Indepe~ Prot~ Other    24     1964
## 7  2008 Never married  21 White Don't ~ Indepe~ Cath~ Not ~    24     1987
## 8  2008 Widowed      71 White Not ap~ Strong~ Prot~ Don'~    24     1937
## 9  2010 Widowed      62 Black Not ap~ Strong~ Prot~ Am b~    24     1948
## 10 2010 Widowed      52 Black Refused Strong~ Prot~ Bapt~    24     1958
## # i 3,308 more rows
```

### 4. Mean TV by religion, sort.

```
gss_cat |> group_by(relig) |>
  summarise(mtv = mean(tvhours, na.rm = TRUE)) |>
  arrange(desc(mtv))
```

```
## # A tibble: 15 x 2
##   relig      mtv
##   <fct>    <dbl>
## 1 Don't know 4.62
## 2 Native american 3.46
## 3 Protestant 3.15
```

##	4 Catholic	2.96
##	5 Inter-nondenominational	2.87
##	6 Christian	2.79
##	7 Other	2.73
##	8 No answer	2.72
##	9 None	2.71
##	10 Jewish	2.52
##	11 Moslem/islam	2.44
##	12 Orthodox-christian	2.42
##	13 Buddhism	2.38
##	14 Hinduism	1.89
##	15 Other eastern	1.67