

Manipulating Factors

Tools for manipulating factors



Set up:

- Install forcats
- Load forcats

General Social Survey

"Since 1972, the General Social Survey (GSS) has provided politicians, policymakers, and scholars with a clear and unbiased perspective on what Americans think and feel about such issues as national spending priorities, crime and punishment, intergroup relations, and confidence in institutions."

<http://gss.norc.org/>

stringsAsFactors woes

- `read.csv()` converts character strings into factors by default

```
glimpse(gss_cat)
```

```
Observations: 21,483
```

```
Variables: 9
```

```
$ year      <int> 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2000, 2...
```

```
$ marital   <fctr> Never married, Divorced, Widowed, Never married, Divorced, Married...
```

```
$ age       <int> 26, 48, 67, 39, 25, 25, 36, 44, 44, 47, 53, 52, 52, 51, 52, 40, 77,...
```

```
$ race      <fctr> White, White, White, White, White, White, White, White, White, Whi...
```

```
$ rincome   <fctr> $8000 to 9999, $8000 to 9999, Not applicable, Not applicable, Not ...
```

```
$ partyid   <fctr> Ind,near rep, Not str republican, Independent, Ind,near rep, Not s...
```

```
$ relig     <fctr> Protestant, Protestant, Protestant, Orthodox-christian, None, Prot...
```

```
$ denom     <fctr> Southern baptist, Baptist-dk which, No denomination, Not applicabl...
```

```
$ tvhours   <int> 12, NA, 2, 4, 1, NA, 3, NA, 0, 3, 2, NA, 1, NA, 1, 7, NA, 3, 3, NA,...
```

Explore factor levels

```
levels(gss_cat$marital)
[1] "No answer"      "Never married" "Separated"      "Divorced"      "Widowed"
[6] "Married"

levels(gss_cat$partyid)
[1] "No answer"      "Don't know"      "Other party"      "Strong republican"
[5] "Not str republican" "Ind,near rep"      "Independent"      "Ind,near dem"
[9] "Not str democrat" "Strong democrat"

levels(gss_cat$relig)
[1] "No answer"      "Don't know"      "Inter-nondenominational"
[4] "Native american" "Christian"      "Orthodox-christian"
[7] "Moslem/islam"    "Other eastern"   "Hinduism"
[10] "Buddhism"        "Other"           "None"
[13] "Jewish"          "Catholic"        "Protestant"
[16] "Not applicable"
```

Modify factor levels

```
gss_cat %>%
  count(partyid)
# A tibble: 10 × 2
  partyid      n
  <fctr> <int>
1   No answer  154
2  Don't know    1
3  Other party  393
4 Strong republican 2314
5 Not str republican 3032
6   Ind,near rep  1791
7   Independent  4119
8   Ind,near dem  2499
9 Not str democrat 3690
10 Strong democrat 3490
```

Modify factor levels

`fct_recode()` allows us to specify new labels for levels

```
gss_cat <-  
  gss_cat %>%  
  mutate(partyid = fct_recode(partyid,  
    "Republican, strong"      = "Strong republican",  
    "Republican, weak"       = "Not str republican",  
    "Independent, near rep"  = "Ind,near rep",  
    "Independent, near dem"  = "Ind,near dem",  
    "Democrat, weak"        = "Not str democrat",  
    "Democrat, strong"      = "Strong democrat"))
```

```
gss_cat %>%
  count(partyid)
# A tibble: 10 × 2
      partyid      n
    <fctr> <int>
1   No answer   154
2  Don't know     1
3  Other party   393
4 Republican, strong 2314
5 Republican, weak  3032
6 Independent, near rep 1791
7   Independent  4119
8 Independent, near dem 2499
9   Democrat, weak  3690
10  Democrat, strong 3490
```


Collapse factor levels

`fct_collapse()` combines levels

```
gss_cat <-  
  gss_cat %>%  
  mutate(partyid = fct_collapse(partyid,  
    other = c("No answer", "Don't know", "Other party"),  
    rep = c("Strong republican", "Not str republican"),  
    ind = c("Ind,near rep", "Independent", "Ind,near dem"),  
    dem = c("Not str democrat", "Strong democrat")  
  ))
```

```
gss_cat %>%  
  count(partyid)
```

```
# A tibble: 4 × 2  
  partyid      n  
  <fctr> <int>  
1   other   548  
2     rep  5346  
3     ind  8409  
4     dem  7180
```

Lump small factor levels together

```
gss_cat %>%
  count(relig) %>%
  arrange(desc(n))
# A tibble: 15 × 2
```

	relig <fctr>	n <int>
1	Protestant	10846
2	Catholic	5124
3	None	3523
4	Christian	689
5	Jewish	388
6	Other	224
7	Buddhism	147
8	Inter-nondenominational	109
9	Moslem/islam	104
10	Orthodox-christian	95
11	No answer	93
12	Hinduism	71
13	Other eastern	32
14	Native american	23
15	Don't know	15

Lump small factor levels together

`fct_lump()` lumps together small groups

```
gss_cat %>%  
  mutate(relig = fct_lump(relig, n = 5)) %>%  
  count(relig) %>%  
  arrange(desc(n))  
# A tibble: 6 × 2  
  relig      n  
  <fctr> <int>  
1 Protestant 10846  
2   Catholic  5124  
3      None   3523  
4      Other    913  
5   Christian   689  
6     Jewish   388
```