team

Michael Egle michaelegle@iastate.edu, John Chandara chandara@iastate.edu

2/19/2020

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggtern)
```

```
## Registered S3 methods overwritten by 'ggtern':
##   method           from
##   +.gg             ggplot2
##   grid.draw.ggplot ggplot2
##   plot.ggplot      ggplot2
##   print.ggplot     ggplot2
```

```
## --
## Remember to cite, run citation(package = 'ggtern') for further info.
## --
```

```
##
## Attaching package: 'ggtern'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, aes, annotate, calc_element, ggplot, ggplot_build,
##     ggplot_gtable, ggplotGrob, ggsave, layer_data, theme, theme_bw,
##     theme_classic, theme_dark, theme_gray, theme_light, theme_linedraw,
##     theme_minimal, theme_void
```

```
install.packages('readxl')
```

```
## Installing package into '/home/johnny/R/x86_64-pc-linux-gnu-library/3.6'
## (as 'lib' is unspecified)
```

```
library(readxl)
dat <- readxl::read_xls('GSS.xls')
str(dat)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    64814 obs. of  10 variables:
##  $ Gss year for this respondent: num  1972 1972 1972 1972 1972 ...
```

```
##  $ General happiness           : chr  "Not too happy" "Not too happy" "Pretty happy" "Not too happy"
##  $ Political party affiliation : chr  "Ind,near dem" "Not str democrat" "Independent" "Not str democ
##  $ Region of residence, age 16 : chr  "Middle atlantic" "E. nor. central" "E. nor. central" "Foreign
##  $ Respondents sex             : chr  "Female" "Male" "Female" "Female" ...
##  $ Rs highest degree           : chr  "Bachelor" "Lt high school" "High school" "Bachelor" ...
##  $ Number of children          : chr  "0" "5" "4" "0" ...
##  $ Marital status              : chr  "Never married" "Married" "Married" "Married" ...
##  $ Respondent id number        : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ Ballot used for interview   : chr  "Not applicable" "Not applicable" "Not applicable" "Not applica
```

Data Cleaning

```
indx <- sapply(dat, is.character)
dat[indx] <- lapply(dat[indx], function(x) as.factor(as.character(x)))
names(dat) <- c('year', 'happiness', 'party', 'residence', 'sex', 'education', 'children', 'marriage',
dat <- droplevels(dat[dat$happiness != 'No answer' & dat$happiness != 'Don\'t know' & dat$happiness !=
dat <- droplevels(dat[dat$party != 'No answer', ])
dat$happiness <- factor(dat$happiness, c('Very happy', 'Pretty happy', 'Not too happy'))
str(dat)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    59709 obs. of  10 variables:
##  $ year     : num  1972 1972 1972 1972 1972 ...
##  $ happiness: Factor w/ 3 levels "Very happy","Pretty happy",..: 3 3 2 3 2 2 3 3 2 2 ...
##  $ party    : Factor w/ 9 levels "Don't know","Ind,near dem",..: 2 5 4 5 8 2 2 2 8 8 ...
##  $ residence: Factor w/ 10 levels "E. nor. central",..: 4 1 1 3 1 1 1 4 10 10 ...
##  $ sex      : Factor w/ 2 levels "Female","Male": 1 2 1 1 1 2 2 2 1 1 ...
##  $ education: Factor w/ 7 levels "Bachelor","Don't know",..: 1 6 4 1 4 4 4 1 4 4 ...
##  $ children : Factor w/ 10 levels "0","1","2","3",..: 1 6 5 1 3 1 3 1 3 5 ...
##  $ marriage : Factor w/ 6 levels "Divorced","Married",..: 3 2 2 2 2 3 1 3 3 2 ...
##  $ id       : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ ballot   : Factor w/ 4 levels "Ballot a","Ballot b",..: 4 4 4 4 4 4 4 4 4 4 ...
```

```
rotatedAxisElementText = function(angle,position='x'){
  angle     = angle[1];
  position  = position[1]
  positions = list(x=0,y=90,top=180,right=270)
  if(!position %in% names(positions))
    stop(sprintf("'position' must be one of [%s]",paste(names(positions),collapse=", ")),call.=FALSE)
  if(!is.numeric(angle))
    stop("'angle' must be numeric",call.=FALSE)
  rads  = (angle - positions[[ position ]])*pi/180
  hjust = 0.5*(1 - sin(rads))
  vjust = 0.5*(1 + cos(rads))
  element_text(angle=angle,vjust=vjust,hjust=hjust)
}
```

Now we have the string columns changed to factors.

```
dat$children <- as.numeric(dat$children)
str(dat)
```

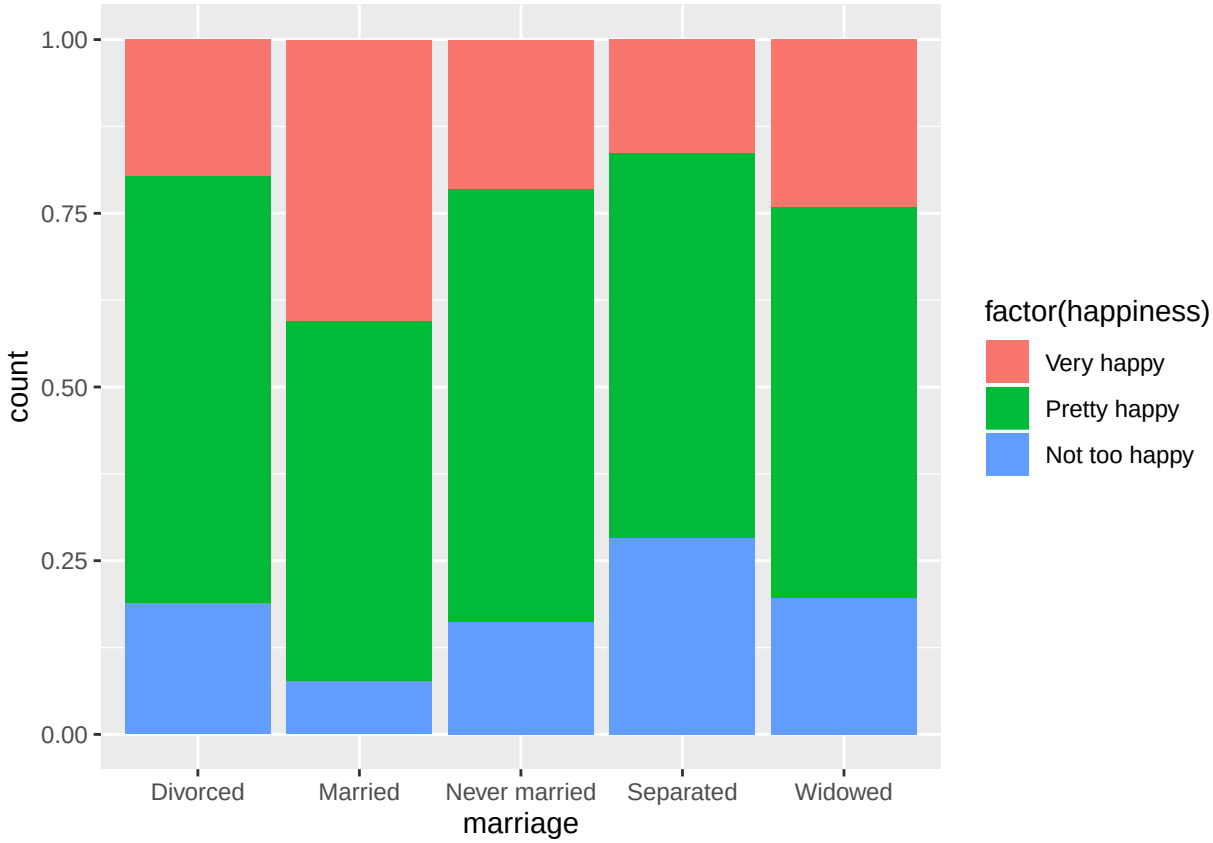
```
## Classes 'tbl_df', 'tbl' and 'data.frame':    59709 obs. of  10 variables:
##  $ year     : num  1972 1972 1972 1972 1972 ...
##  $ happiness: Factor w/ 3 levels "Very happy","Pretty happy",..: 3 3 2 3 2 2 3 3 2 2 ...
##  $ party    : Factor w/ 9 levels "Don't know","Ind,near dem",..: 2 5 4 5 8 2 2 2 8 8 ...
```

```
##  $ residence: Factor w/ 10 levels "E. nor. central",..: 4 1 1 3 1 1 1 4 10 10 ...
##  $ sex      : Factor w/ 2 levels "Female","Male": 1 2 1 1 1 2 2 2 1 1 ...
##  $ education: Factor w/ 7 levels "Bachelor","Don't know",..: 1 6 4 1 4 4 4 1 4 4 ...
##  $ children : num  1 6 5 1 3 1 3 1 3 5 ...
##  $ marriage : Factor w/ 6 levels "Divorced","Married",..: 3 2 2 2 2 3 1 3 3 2 ...
##  $ id       : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ ballot   : Factor w/ 4 levels "Ballot a","Ballot b",..: 4 4 4 4 4 4 4 4 4 4 ...
```

Exploration

How does the happiness of a respondent relate to the marriage status?

```
dat %>%
  filter(marriage != "No answer", marriage != "NA") %>%
  ggplot(aes(x = marriage, fill = factor(happiness))) +
  geom_bar(position = "fill")
```

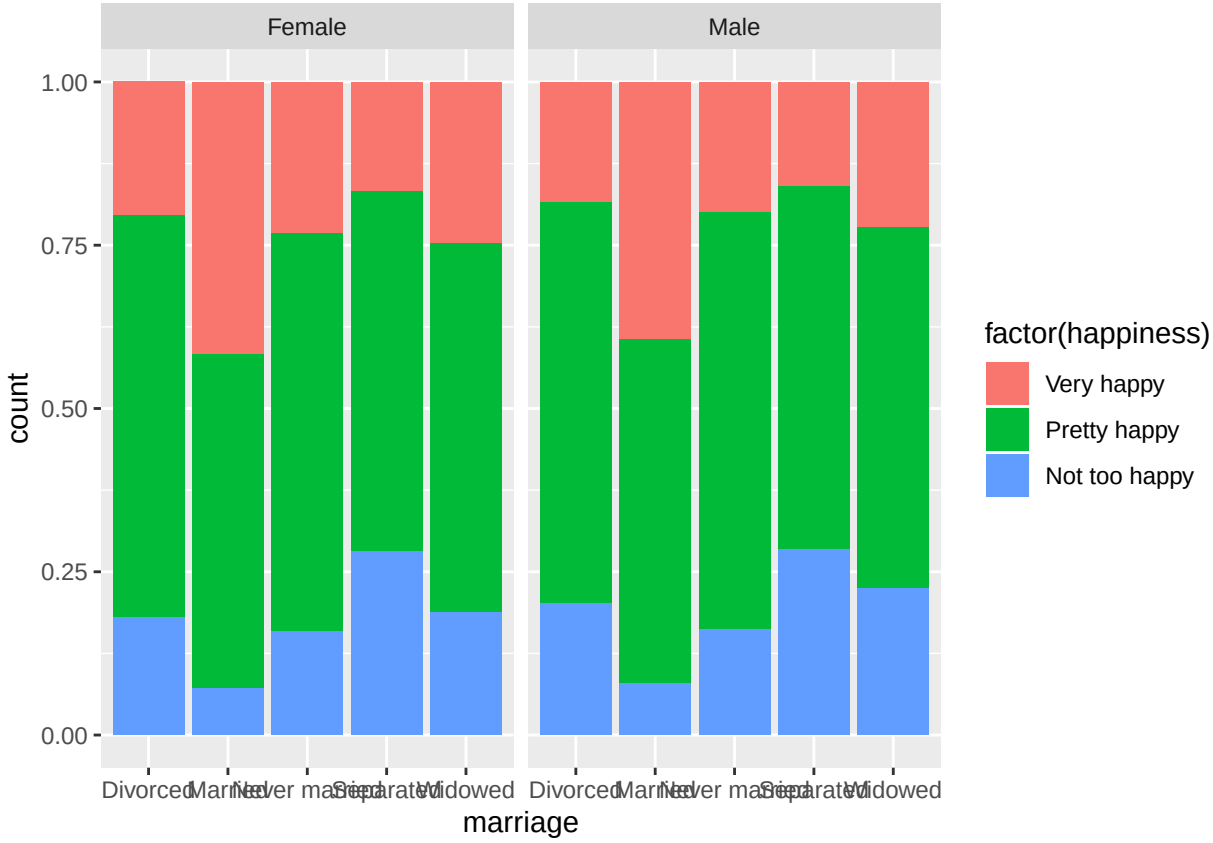


People who are currently married have the highest proportion of being "Very happy" and "Pretty happy". Married respondents also had the lowest proportion of "Not too happy" responses. Separated and divorced respondents had the highest proportion of "Not too happy" responses. There appears to be a positive relationship between marital status and happiness of the respondent.

Does the sex of the respondent affect the relationship you found in Q1?

```
dat %>%
  filter(marriage != "No answer", marriage != "NA") %>%
```

```
  ggplot(aes(x = marriage, fill = factor(happiness))) +
  geom_bar(position = "fill") +
  facet_grid(. ~ sex)
```

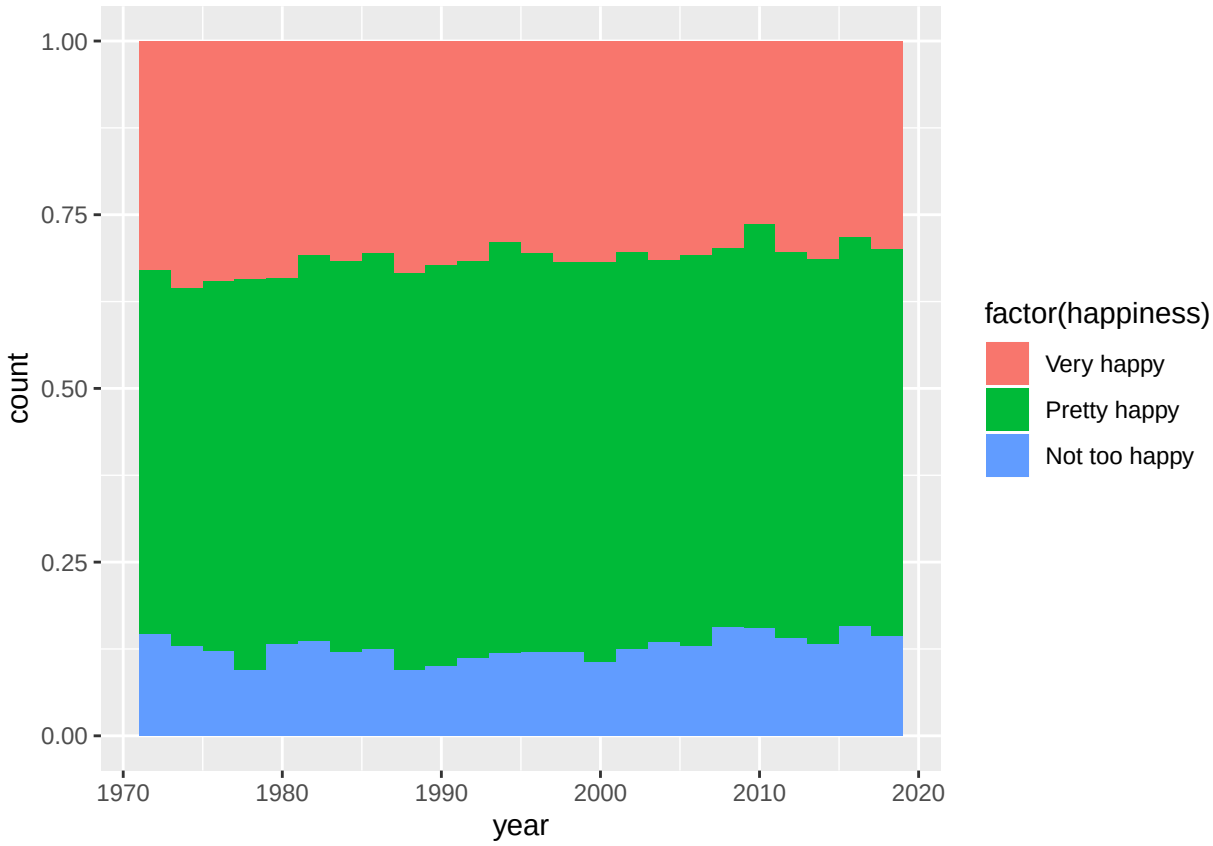


The graphs are nearly identical between Males and Females, so I do not believe the sex of the respondent has an effect on the relationship between marital status and happiness.

How has the distribution of happiness changed over the years?

```
dat %>%
  filter(year != "NA") %>%
  ggplot(aes(x = year, fill = factor(happiness))) +
  geom_bar(position = "fill", binwidth = 2)
```

```
## Warning: `geom_bar()` no longer has a `binwidth` parameter. Please use
## `geom_histogram()` instead.
```



There isn't a ton of variance but it does appear that there's been a slight incline in "Not too happy" responses over the last 30 years.

Have political views become more polarized (i.e. more "Strong" republican/democrats in recent years) over time?

```
unique(factor(dat$party))
```

```
## [1] Ind,near dem       Not str democrat   Independent        Strong democrat
## [5] Not str republican Ind,near rep       Strong republican  Other party
## [9] Don't know
## 9 Levels: Don't know Ind,near dem Ind,near rep ... Strong republican
```

Let's relevel these quick to make them go in order in our graphic.

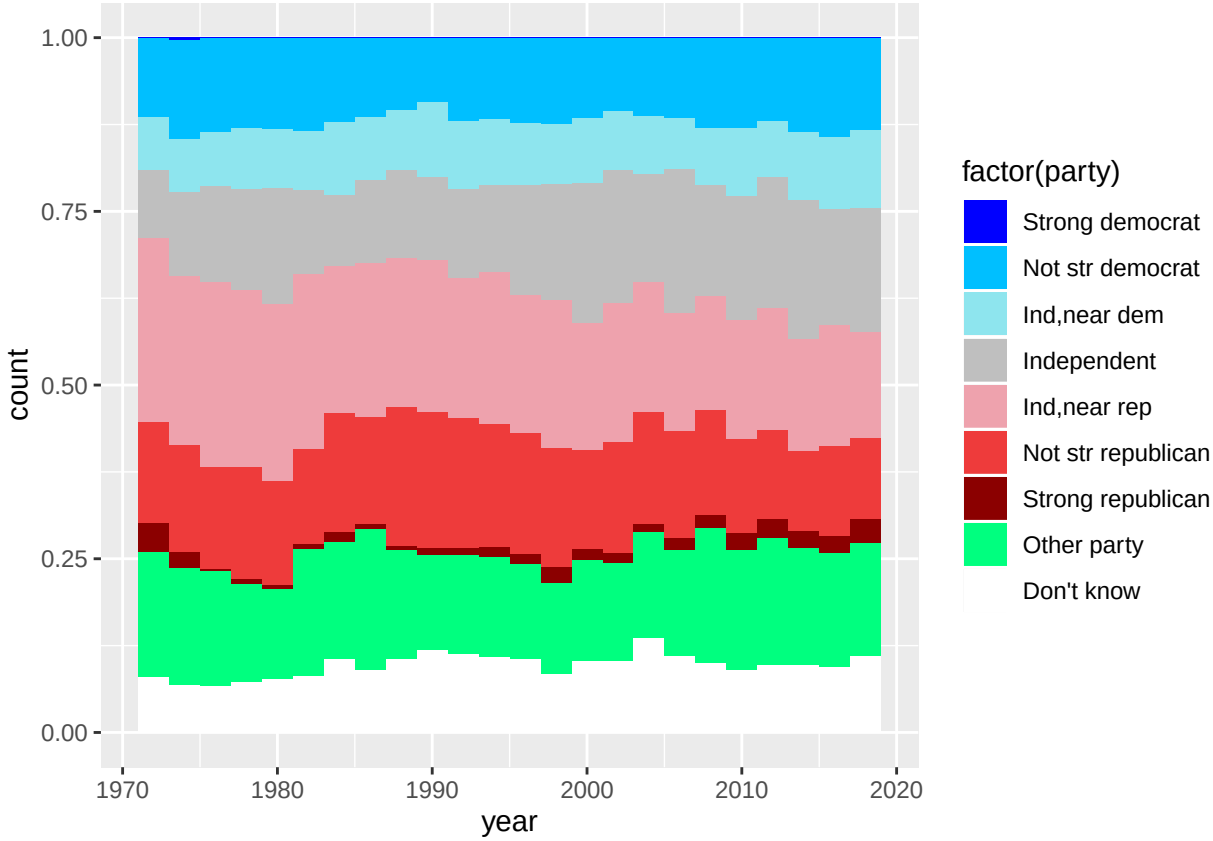
```
party_levels <- c("Strong democrat", "Not str democrat", "Ind,near dem",
                  "Independent", "Ind,near rep", "Not str republican",
                  "Strong republican", "Other party", "Don't know")
levels(dat$party) <- party_levels
levels(dat$party)
```

```
## [1] "Strong democrat"    "Not str democrat"   "Ind,near dem"
## [4] "Independent"        "Ind,near rep"       "Not str republican"
## [7] "Strong republican"  "Other party"        "Don't know"
```

```
dat %>%
  filter(party != "NA") %>%
  ggplot(aes(x = year, fill = factor(party))) +
```

```
  geom_bar(position = "fill", binwidth = 2) +
  scale_fill_manual(values = c("blue", "deepskyblue1", "cadetblue2",
                               "gray75", "lightpink2", "brown2",
                               "red4", "springgreen", "white"),
                    limits = party_levels)
```

```
## Warning: `geom_bar()` no longer has a `binwidth` parameter. Please use
## `geom_histogram()` instead.
```

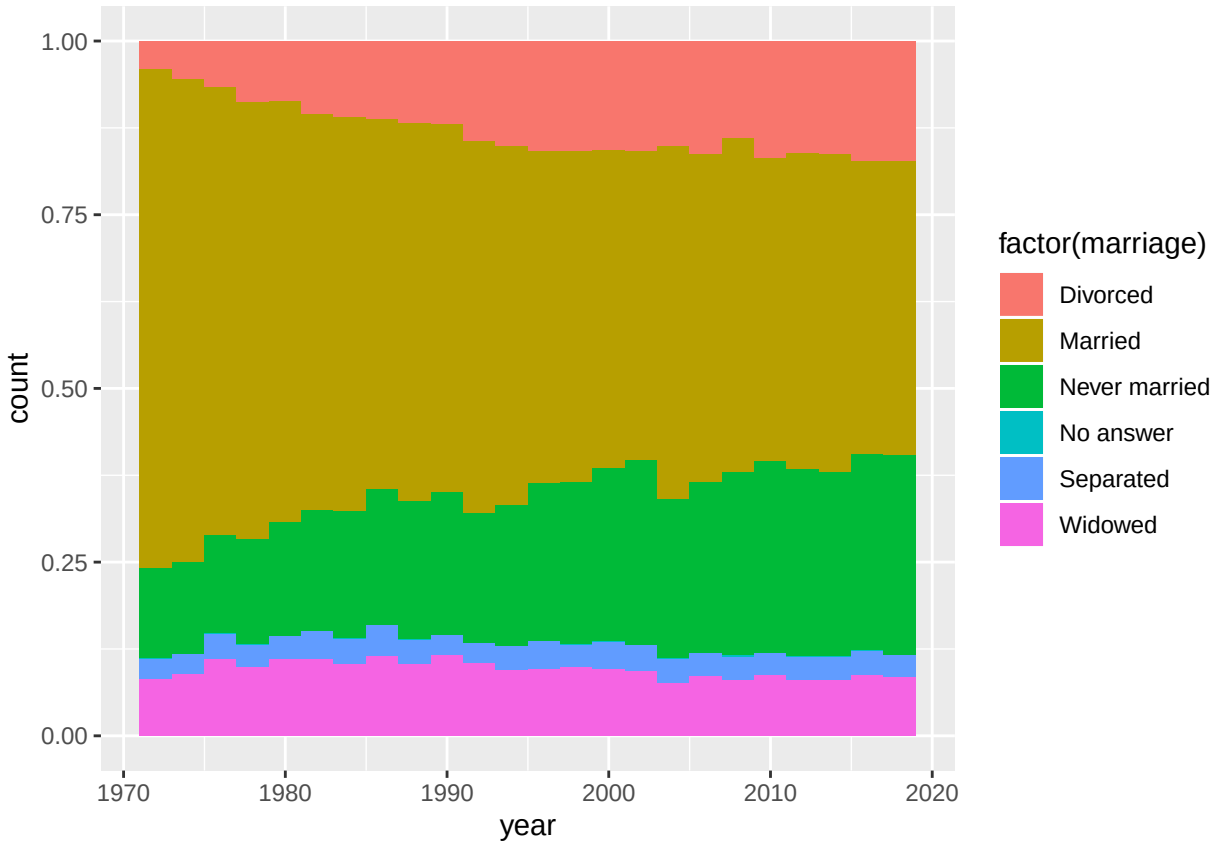


The answer isn't entirely clear. Depending on how you define "political polarization", it could be yes or no. There's been an overall decrease in republican identifying respondents since 1990 but a reasonable increase in "Strong" republican identifying respondents in that same period. Meanwhile there has been a slight increase in democrats. So this could be considered a "polarization". On the other hand. There has been an overall decrease in republicans and increase in "other party" / "don't know" respondents recently.

How has the proportion of marital statuses change over time?

```
dat %>%
  filter(year != "NA") %>%
  ggplot(aes(x = year, fill = factor(marriage))) +
  geom_bar(position = "fill", binwidth = 2)
```

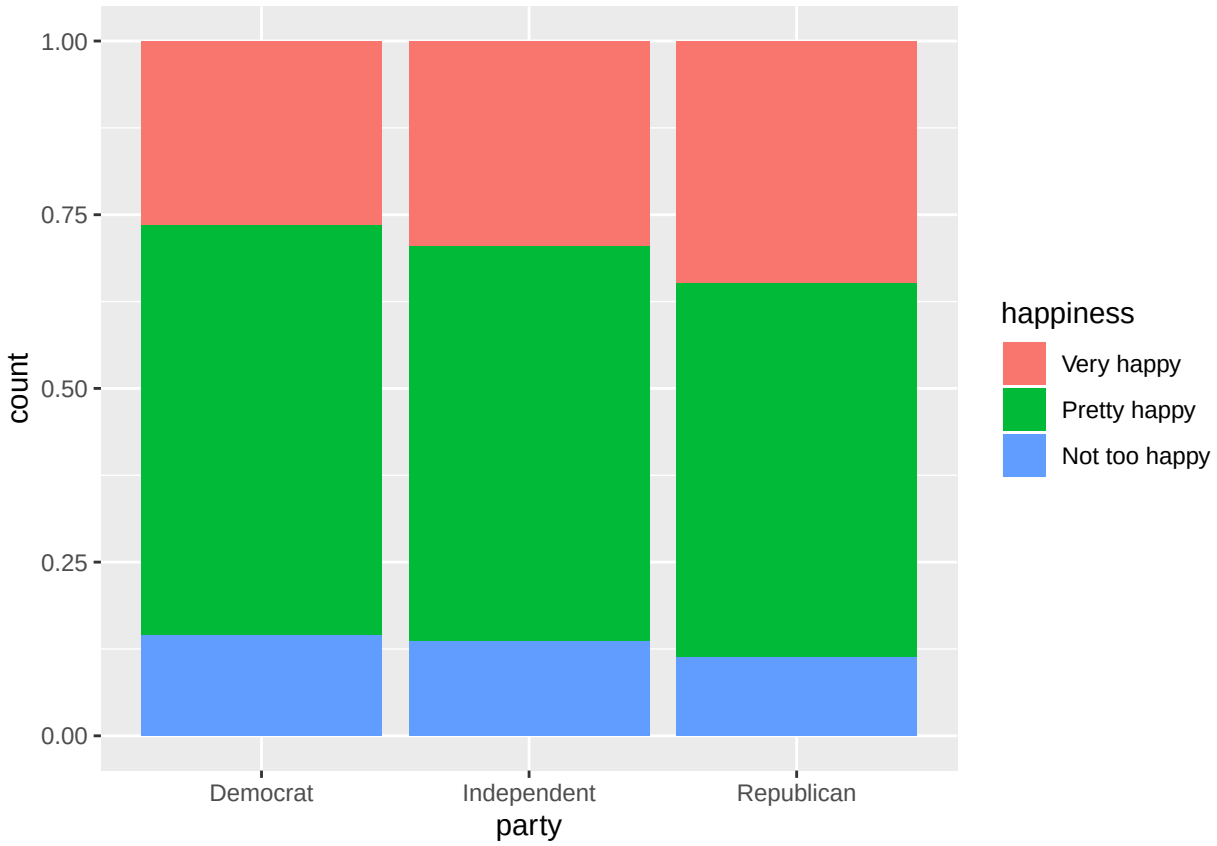
```
## Warning: `geom_bar()` no longer has a `binwidth` parameter. Please use
## `geom_histogram()` instead.
```



There are a few pretty noticeable trends in the data. The proportion of married respondents has decreased a lot from 1970 to the present. There's also been a big increase in respondents who were divorced or never married.

How does the happiness of a respondent relate to the political party affiliation?

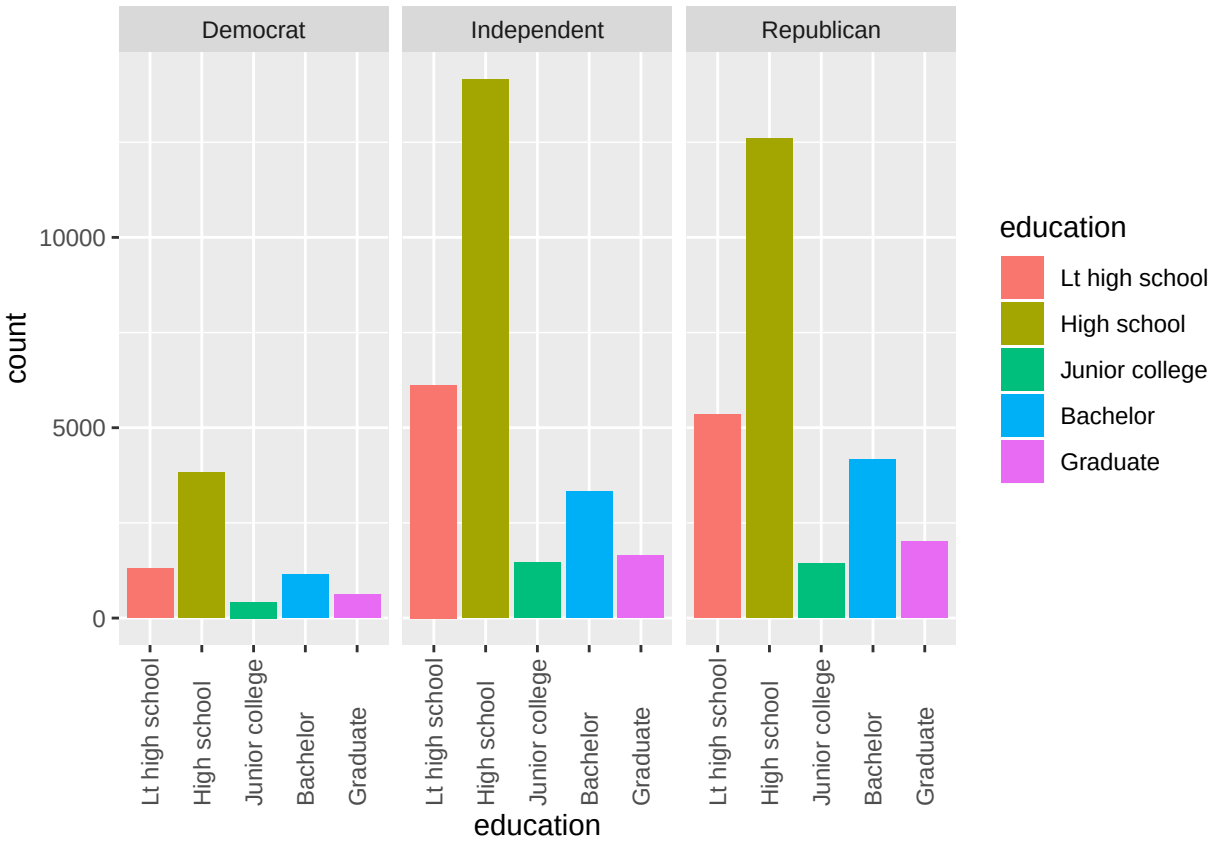
```
dat2 <- dat

parties <- levels(dat2$party)
levels(dat2$party)[levels(dat2$party) == 'Not str democrat' | levels(dat2$party) == 'Strong democrat'

levels(dat2$party)[levels(dat2$party) == 'Not str republican' | levels(dat2$party) == 'Strong republi

levels(dat2$party)[levels(dat2$party) == 'Don\'t know' | levels(dat2$party) == 'Other party'] <- 'Repu

levels(dat2$party)[levels(dat2$party) == 'Independent' | levels(dat2$party) == 'Ind,near dem' | levels

dat2 %>%
  ggplot(aes(x = party, fill = happiness)) + geom_bar(position = 'fill')
```



Based on the results, republicans appear to be much more happier than independent and democratic respondents by a well margin.

Is political affiliation affected by education?

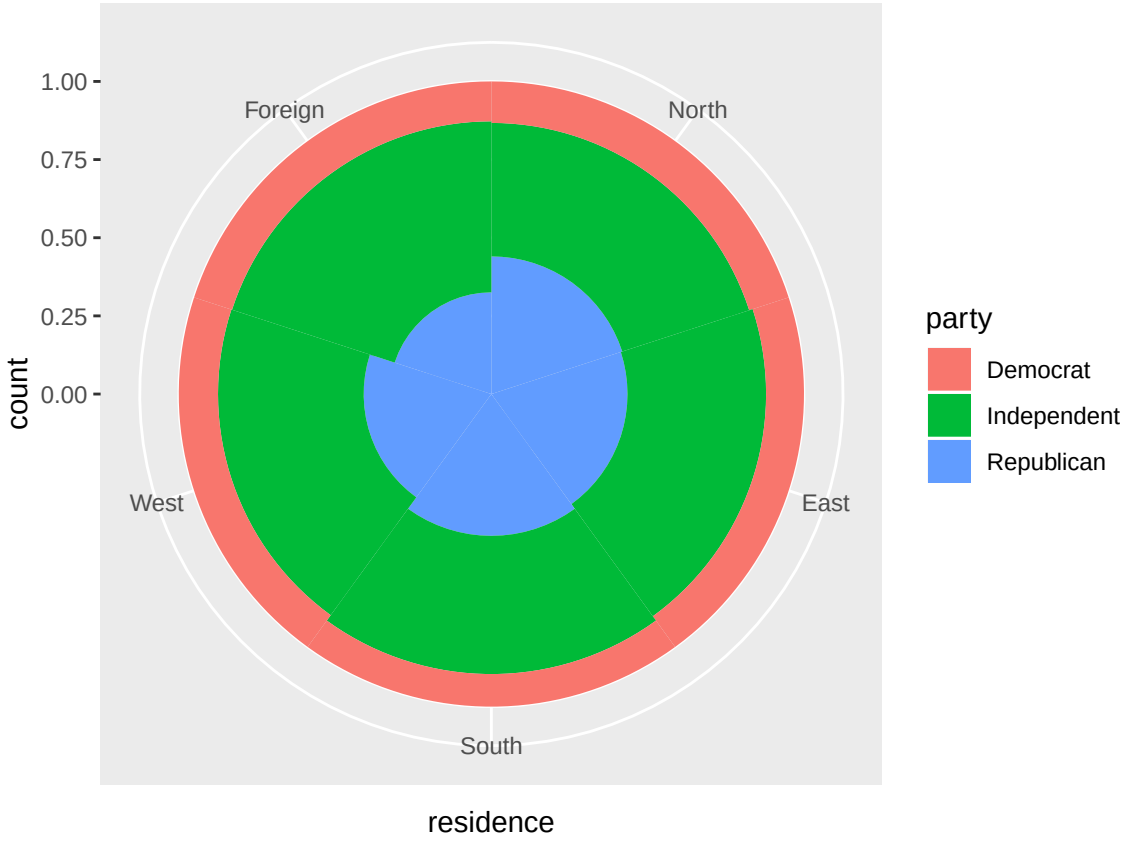
```
dat2$education <- factor(dat2$education, c('No answer', 'Don\'t know','Lt high school', 'High school'

filtered <- droplevels(dat2[dat2$education != 'No answer' & dat2$education != 'Don\'t know', ])

filtered %>%
  ggplot(aes(x = education, fill = education)) + geom_bar() + facet_wrap(~party) + theme(axis.text.x =
```



From this data, we can conclude that respondent with the highest education level in junior college typically swing twords independent or democratic ideologies. Time afterwards, their ideology may swing a bit republican however, this isn't strong enough as a small percentage more of full graduates are democrats.

How does the political affiliation affect place of residence?

```
levels(dat2$residence)[levels(dat2$residence) == 'E. nor. central' | levels(dat2$residence) == 'W. no

levels(dat2$residence)[levels(dat2$residence) == 'New england' | levels(dat2$residence) == 'Middle at

levels(dat2$residence)[levels(dat2$residence) == 'E. sou. central' | levels(dat2$residence) == 'W. so

levels(dat2$residence)[levels(dat2$residence) == 'Mountain' | levels(dat2$residence) == 'Pacific' | l

dat2$residence <- factor(dat2$residence, c('North', 'East', 'South', 'West', 'Foreign'))

dat2 %>%
  ggplot(aes(x = residence, fill = party)) + geom_bar(position = 'fill', width = 1) + coord_polar()
```



For the most part, all regions of the United States appear to be in quite equal in representation. Though, we see a bit more Republican respondents in the North and West. For democrats, we see a magnitude more in the south. As for Foreign respondent, there appears to be less republican-centered respondent.