hw4-Deep-Bhalodia

Deep Bhalodia

2/21/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Loading of required packages

```
library(ggplot2)
library(dplyr)
library(rlang)
library(readr)
library(tidyverse)
```

Part A

```
crdc<-read_csv("/Users/deep/downloads/CRDC 2015-16 School Data.csv", guess_max = 15000, na=c("-2", "-5"
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   LEA_STATE = col_character(),
##   LEA_STATE_NAME = col_character(),
##   LEAID = col_character(),
##   LEA_NAME = col_character(),
##   SCH_NAME = col_character(),
##   COMBOKEY = col_character(),
##   JJ = col_character(),
##   SCH_GRADE_PS = col_character(),
##   SCH_GRADE_KG = col_character(),
##   SCH_GRADE_G01 = col_character(),
##   SCH_GRADE_G02 = col_character(),
##   SCH_GRADE_G03 = col_character(),
##   SCH_GRADE_G04 = col_character(),
##   SCH_GRADE_G05 = col_character(),
##   SCH_GRADE_G06 = col_character(),
##   SCH_GRADE_G07 = col_character(),
##   SCH_GRADE_G08 = col_character(),
##   SCH_GRADE_G09 = col_character(),
##   SCH_GRADE_G10 = col_character(),
##   SCH_GRADE_G11 = col_character()
##   # ... with 45 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
echo = FALSE

crdc
```

```
## # A tibble: 96,360 x 1,836
##    LEA_STATE LEA_STATE_NAME LEAID LEA_NAME SCHID SCH_NAME COMBOKEY JJ
##    <chr>     <chr>          <chr> <chr>    <dbl> <chr>    <chr>    <chr>
##  1 AL        ALABAMA        1000~ Alabama~  1705 Wallace~ 1000020~ Yes
##  2 AL        ALABAMA        1000~ Alabama~  1706 McNeel ~ 1000020~ Yes
##  3 AL        ALABAMA        1000~ Alabama~  1876 Alabama~ 1000020~ No
##  4 AL        ALABAMA        1000~ Alabama~ 99995 AUTAUGA~ 1000029~ Yes
##  5 AL        ALABAMA        1000~ Albertv~   870 Albertv~ 1000050~ No
##  6 AL        ALABAMA        1000~ Albertv~   871 Albertv~ 1000050~ No
##  7 AL        ALABAMA        1000~ Albertv~   879 Evans E~ 1000050~ No
##  8 AL        ALABAMA        1000~ Albertv~   889 Albertv~ 1000050~ No
##  9 AL        ALABAMA        1000~ Albertv~  1616 Big Spr~ 1000050~ No
## 10 AL        ALABAMA        1000~ Albertv~  2150 Albertv~ 1000050~ No
## # ... with 96,350 more rows, and 1,828 more variables: SCH_GRADE_PS <chr>,
## #   SCH_GRADE_KG <chr>, SCH_GRADE_G01 <chr>, SCH_GRADE_G02 <chr>,
## #   SCH_GRADE_G03 <chr>, SCH_GRADE_G04 <chr>, SCH_GRADE_G05 <chr>,
## #   SCH_GRADE_G06 <chr>, SCH_GRADE_G07 <chr>, SCH_GRADE_G08 <chr>,
## #   SCH_GRADE_G09 <chr>, SCH_GRADE_G10 <chr>, SCH_GRADE_G11 <chr>,
## #   SCH_GRADE_G12 <chr>, SCH_GRADE_UG <chr>, SCH_UGDETAIL_ES <chr>,
## #   SCH_UGDETAIL_MS <chr>, SCH_UGDETAIL_HS <chr>, SCH_STATUS_SPED <chr>,
## #   SCH_STATUS_MAGNET <chr>, SCH_STATUS_CHARTER <chr>,
## #   SCH_STATUS_ALT <chr>, SCH_MAGNETDETAIL <chr>, SCH_ALTFOCUS <chr>,
## #   SCH_PSENR_NONIDEA_A3 <chr>, SCH_PSENR_NONIDEA_A4 <chr>,
## #   SCH_PSENR_NONIDEA_A5 <chr>, SCH_PSENR_HI_M <dbl>,
## #   SCH_PSENR_HI_F <dbl>, SCH_PSENR_AM_M <dbl>, SCH_PSENR_AM_F <dbl>,
## #   SCH_PSENR_AS_M <dbl>, SCH_PSENR_AS_F <dbl>, SCH_PSENR_HP_M <dbl>,
## #   SCH_PSENR_HP_F <dbl>, SCH_PSENR_BL_M <dbl>, SCH_PSENR_BL_F <dbl>,
## #   SCH_PSENR_WH_M <dbl>, SCH_PSENR_WH_F <dbl>, SCH_PSENR_TR_M <dbl>,
## #   SCH_PSENR_TR_F <dbl>, TOT_PSENR_M <dbl>, TOT_PSENR_F <dbl>,
## #   SCH_PSENR_LEP_M <dbl>, SCH_PSENR_LEP_F <dbl>, SCH_PSENR_IDEA_M <dbl>,
## #   SCH_PSENR_IDEA_F <dbl>, SCH_ENR_HI_M <dbl>, SCH_ENR_HI_F <dbl>,
## #   SCH_ENR_AM_M <dbl>, SCH_ENR_AM_F <dbl>, SCH_ENR_AS_M <dbl>,
## #   SCH_ENR_AS_F <dbl>, SCH_ENR_HP_M <dbl>, SCH_ENR_HP_F <dbl>,
## #   SCH_ENR_BL_M <dbl>, SCH_ENR_BL_F <dbl>, SCH_ENR_WH_M <dbl>,
## #   SCH_ENR_WH_F <dbl>, SCH_ENR_TR_M <dbl>, SCH_ENR_TR_F <dbl>,
## #   TOT_ENR_M <dbl>, TOT_ENR_F <dbl>, SCH_ENR_LEP_M <dbl>,
## #   SCH_ENR_LEP_F <dbl>, SCH_ENR_504_M <dbl>, SCH_ENR_504_F <dbl>,
## #   SCH_ENR_IDEA_M <dbl>, SCH_ENR_IDEA_F <dbl>, SCH_LEPENR_HI_M <dbl>,
## #   SCH_LEPENR_HI_F <dbl>, SCH_LEPENR_AM_M <dbl>, SCH_LEPENR_AM_F <dbl>,
## #   SCH_LEPENR_AS_M <dbl>, SCH_LEPENR_AS_F <dbl>, SCH_LEPENR_HP_M <dbl>,
## #   SCH_LEPENR_HP_F <dbl>, SCH_LEPENR_BL_M <dbl>, SCH_LEPENR_BL_F <dbl>,
## #   SCH_LEPENR_WH_M <dbl>, SCH_LEPENR_WH_F <dbl>, SCH_LEPENR_TR_M <dbl>,
## #   SCH_LEPENR_TR_F <dbl>, TOT_LEPENR_M <dbl>, TOT_LEPENR_F <dbl>,
## #   SCH_LEPPROGENR_HI_M <dbl>, SCH_LEPPROGENR_HI_F <dbl>,
## #   SCH_LEPPROGENR_AM_M <dbl>, SCH_LEPPROGENR_AM_F <dbl>,
## #   SCH_LEPPROGENR_AS_M <dbl>, SCH_LEPPROGENR_AS_F <dbl>,
## #   SCH_LEPPROGENR_HP_M <dbl>, SCH_LEPPROGENR_HP_F <dbl>,
## #   SCH_LEPPROGENR_BL_M <dbl>, SCH_LEPPROGENR_BL_F <dbl>,
## #   SCH_LEPPROGENR_WH_M <dbl>, SCH_LEPPROGENR_WH_F <dbl>,
```

```
## #   SCH_LEPPROGENR_TR_M <dbl>, SCH_LEPPROGENR_TR_F <dbl>,
## #   TOT_LEPPROGENR_M <dbl>, ...
```

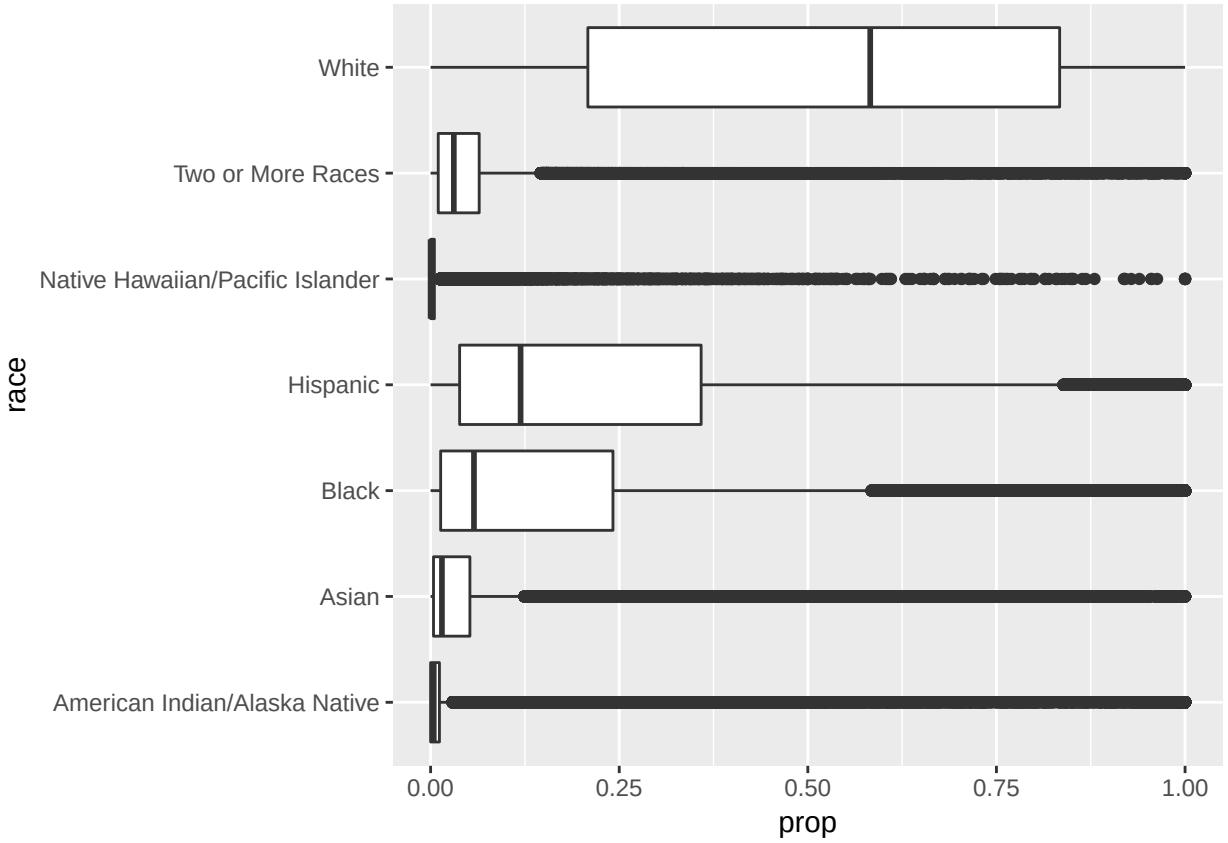
Problem 1

```
crdc_race <- crdc %>%select(SCH_NAME, SCH_ENR_HI_M:SCH_ENR_TR_F) %>%
  gather(key=racegender, value=enrollment,
  SCH_ENR_HI_M:SCH_ENR_TR_F) %>%
  separate(racegender, into=c("race", "gender"),sep=-1) %>%
  mutate(race=recode(race,SCH_ENR_HI_="Hispanic",SCH_ENR_AM_="American Indian/Alaska Native",
  SCH_ENR_AS_="Asian",SCH_ENR_HP_="Native Hawaiian/Pacific Islander",SCH_ENR_BL_="Black",
  SCH_ENR_WH_="White",SCH_ENR_TR_="Two or More Races")) %>%
  group_by(SCH_NAME, race) %>%
  summarize(enrollment = sum(enrollment, na.rm=TRUE)) %>%
  left_join(select(crdc, SCH_NAME, TOT_ENR_M, TOT_ENR_F)) %>%
  mutate(prop = enrollment / (TOT_ENR_M + TOT_ENR_F)) %>%
  filter(prop <= 1)
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
## Joining, by = "SCH_NAME"
```

```
crdc_race %>%ggplot() +
  geom_boxplot(aes(x=race, y=prop)) +
  coord_flip()
```



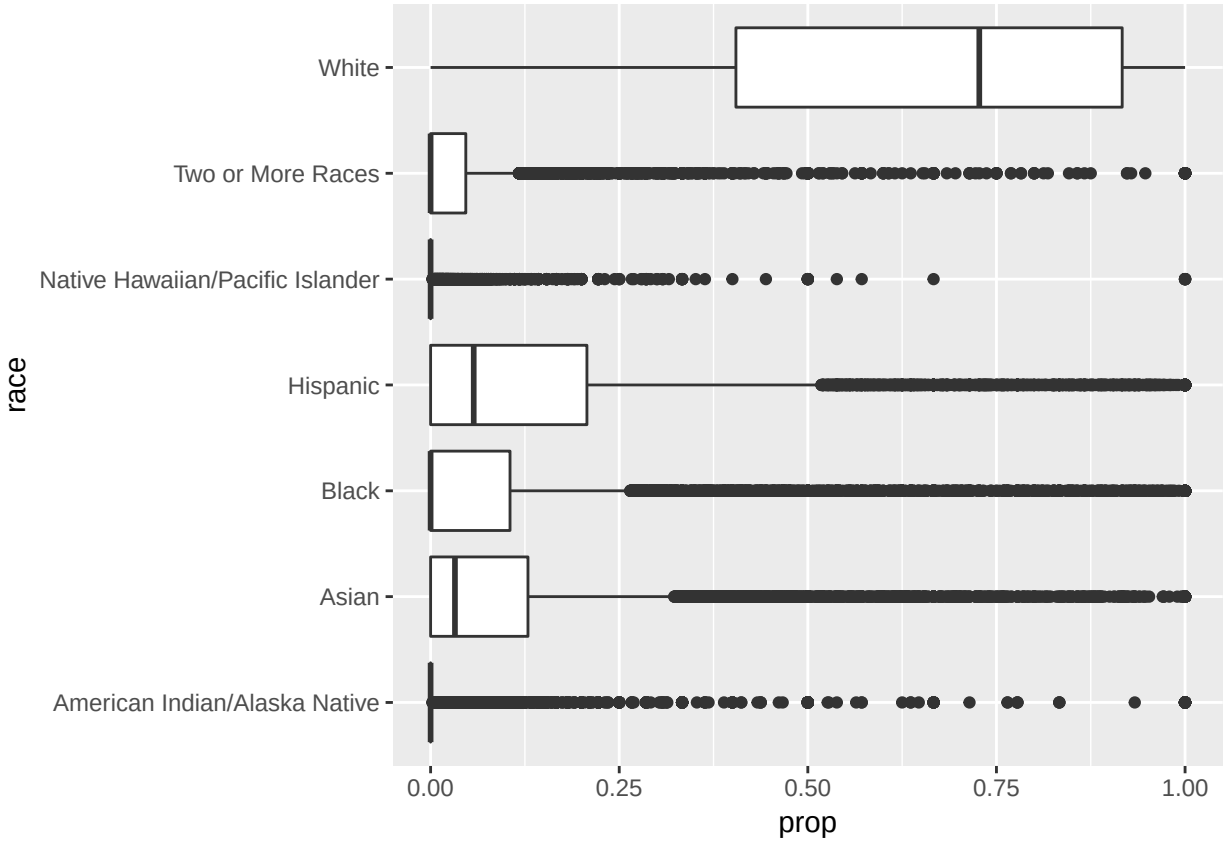
Problem 2

For each school, calculate the proportions of students of each race enrolled in a Calculus class (among students enrolled in calculus classes at each school). (It may be helpful to filter out schools without data for Calculus class enrollment.) Then create side-by-side boxplots showing the distributions of these proportions for each race.

```
crdc_calc <- crdc %>%filter(!is.na(TOT_MATHENR_CALC_M), !is.na(TOT_MATHENR_CALC_F)) %>%
  select(SCH_NAME, SCH_MATHENR_CALC_HI_M:SCH_MATHENR_CALC_TR_F) %>%
  gather(key=racegender, value=calc_enrollment,SCH_MATHENR_CALC_HI_M:SCH_MATHENR_CALC_TR_F) %>%
  separate(racegender, into=c("race", "gender"),sep=-1) %>%
  mutate(race=recode(race,SCH_MATHENR_CALC_HI_="Hispanic",SCH_MATHENR_CALC_AM_="American Indian/Alaska
  group_by(SCH_NAME, race) %>%
  summarize(calc_enrollment = sum(calc_enrollment, na.rm=TRUE)) %>%
  left_join(select(crdc, SCH_NAME, TOT_MATHENR_CALC_M, TOT_MATHENR_CALC_F)) %>%
  mutate(prop = calc_enrollment / (TOT_MATHENR_CALC_M + TOT_MATHENR_CALC_F)) %>%
  filter(prop <= 1)
```

```
## Joining, by = "SCH_NAME"
```

```
crdc_calc %>%ggplot() +
  geom_boxplot(aes(x=race, y=prop)) +
  coord_flip()
```



Comments

The relative proportions of Hispanic, Black, and multiracial students drop dramatically in the bar plot for Calculus enrollment, while the relative proportions of White and Asian students increases between the two plots.

There are two few American Indian / Alaska Native and Native Hawaiian / Pacific Islander students to be able to discern changes in those sub-populations from this plot alone

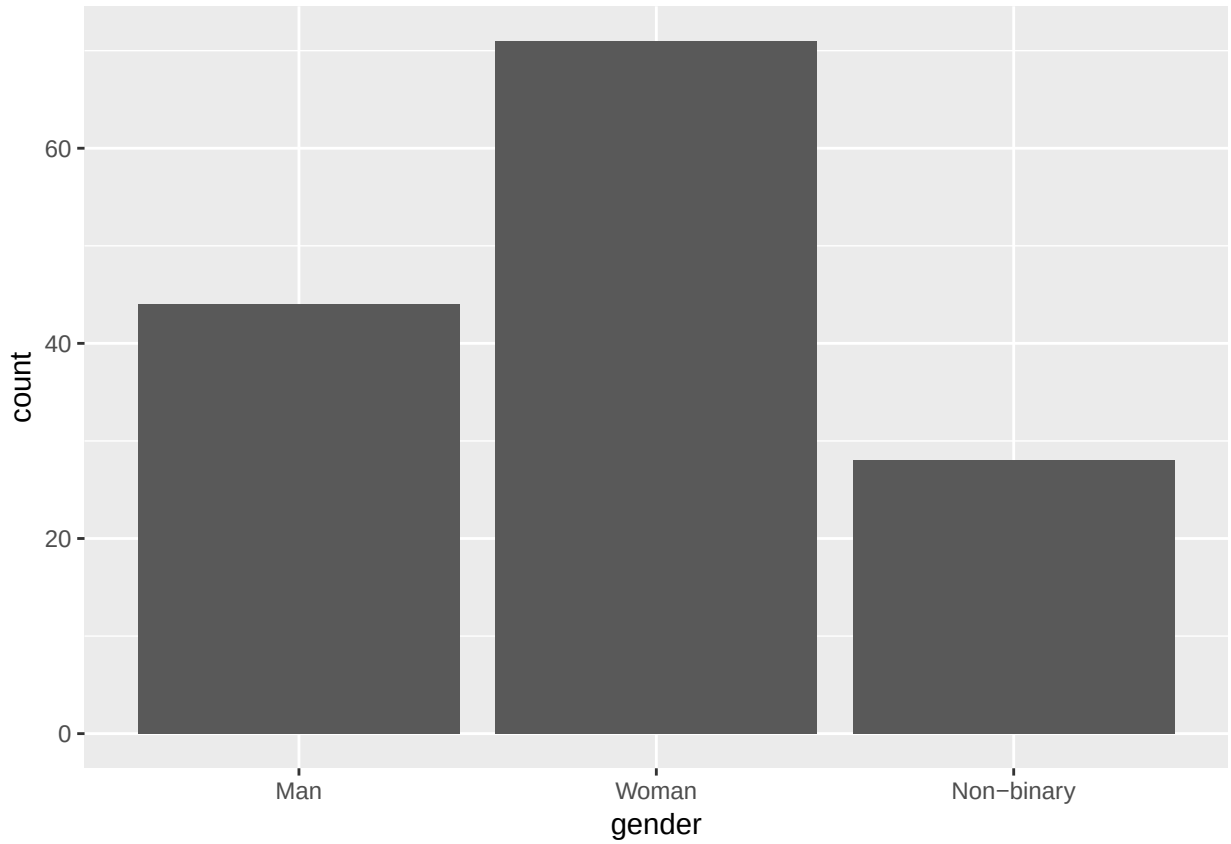
PART B

```
load(file = "31721-0001-Data.rda")

this <- as_tibble(da31721.0001) %>%
  transmute(id=RESPKEY,
  gender=Q6,
  asab=Q5,
  race_black=D9_1,
  race_white=D9_2,
  race_latinx=D9_3,
  race_native=D9_4,
  homeless_ever=Q88,
  fam_support=Q119,
  suicide_thoughts=Q131,
  suicide_attempt=Q133)
```

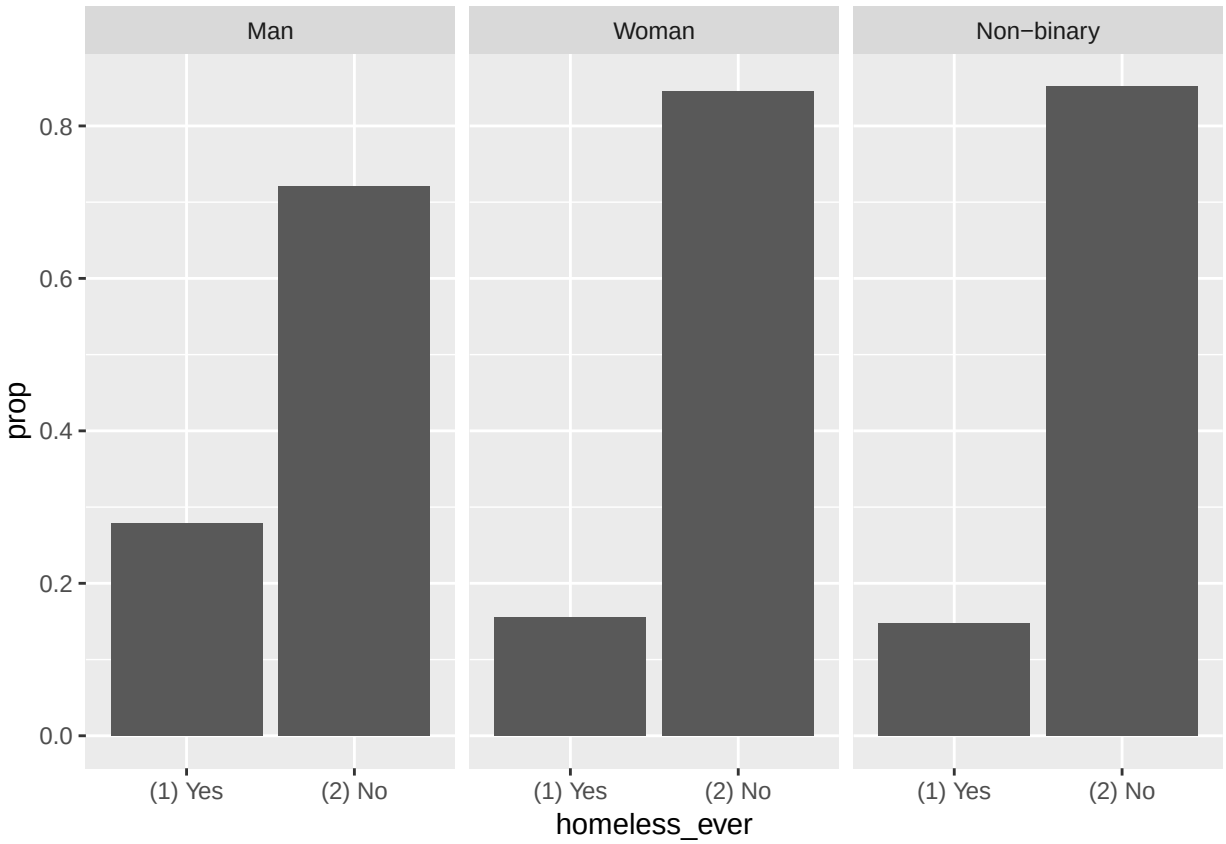
Problem 3

Transform the data to include only 3 gender categories for trans men, trans women, and non-binary participants.

```
this_recode <- this %>%mutate(gender = recode(gender,`(1) Man` = "Man",`(2) Woman` =  "Woman",`(4) Andr
  filter((gender == "Woman" & asab == "(1) Male") |(gender == "Man" & asab == "(2) Female") |gender ==

ggplot(this_recode) + geom_bar(aes(x=gender))
```




```
this_recode %>%
  filter(!is.na(homeless_ever)) %>%
  ggplot() +
  geom_bar(aes(x=homeless_ever,y=..prop..,group=1)) +
  facet_wrap(~gender)
```



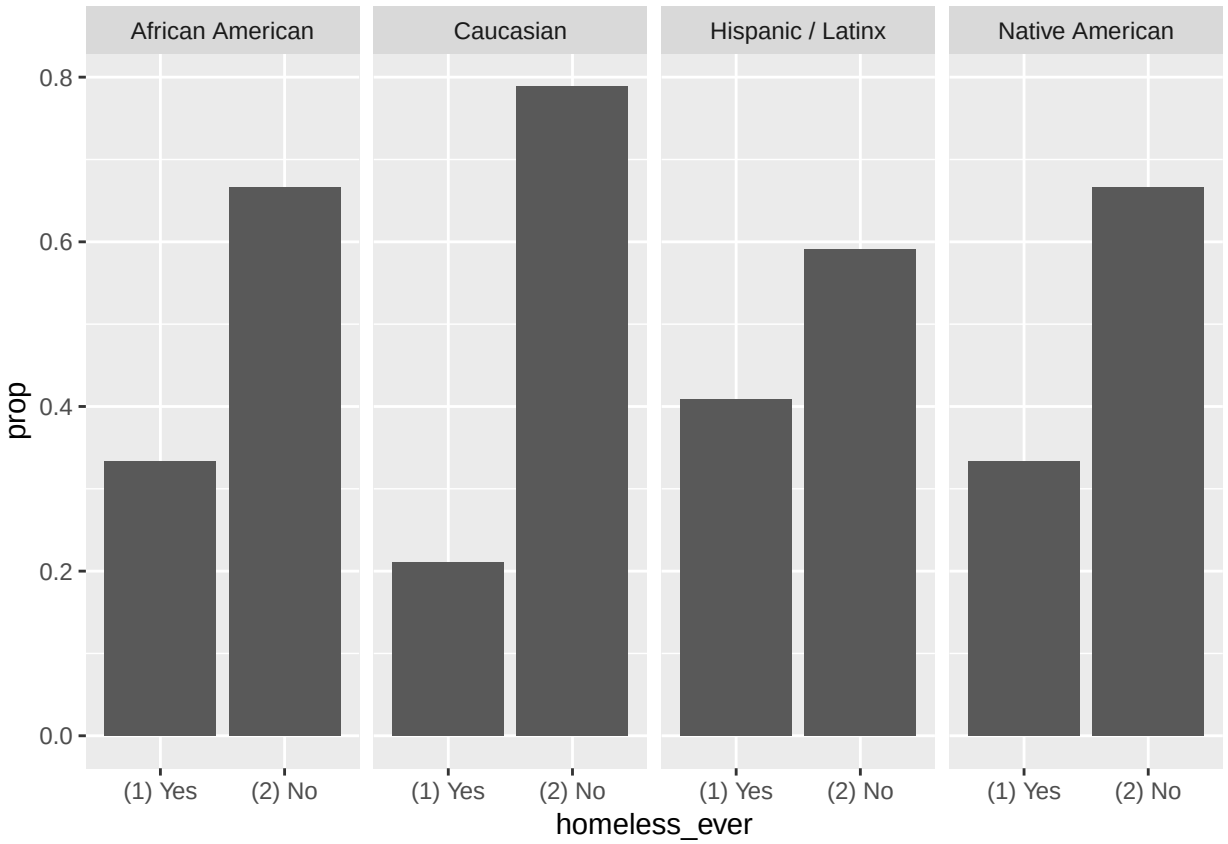
Comments

Trans women, trans men, and non-binary people in this survey have all experienced homelessness at a greater lifetime rate than the U.S. general population.

Problem 4

Using the full dataset again, transform the dataset to have a column for race indicating the race of the participant.

```
this %>%
  gather(key = race,value = race_selected,race_black:race_native) %>%
  filter(race_selected == "(1) Selected") %>%
  select(-race_selected) %>%
  mutate(race = recode(race,race_black = "African American",race_white = "Caucasian",race_latinx = "Hisp
  filter(!is.na(homeless_ever)) %>%
  ggplot() +
  geom_bar(aes(x=homeless_ever, y=..prop.., group=race)) +
  facet_grid(~race)
```



Comments Note that some participants identify as multiple races, so they may appear in multiple facets of the above bar plots (but only once per subplot).

Participants of all races in this survey have experienced homelessness at a greater lifetime rate than the U.S. general population.

Problem 5

Using the full dataset, calculate the total proportion of participants who have attempted suicide in the Virginia THIS survey.

```
this %>%
summarise(suicide=sum(suicide_attempt == "(1) Yes", na.rm=TRUE) / n())
```

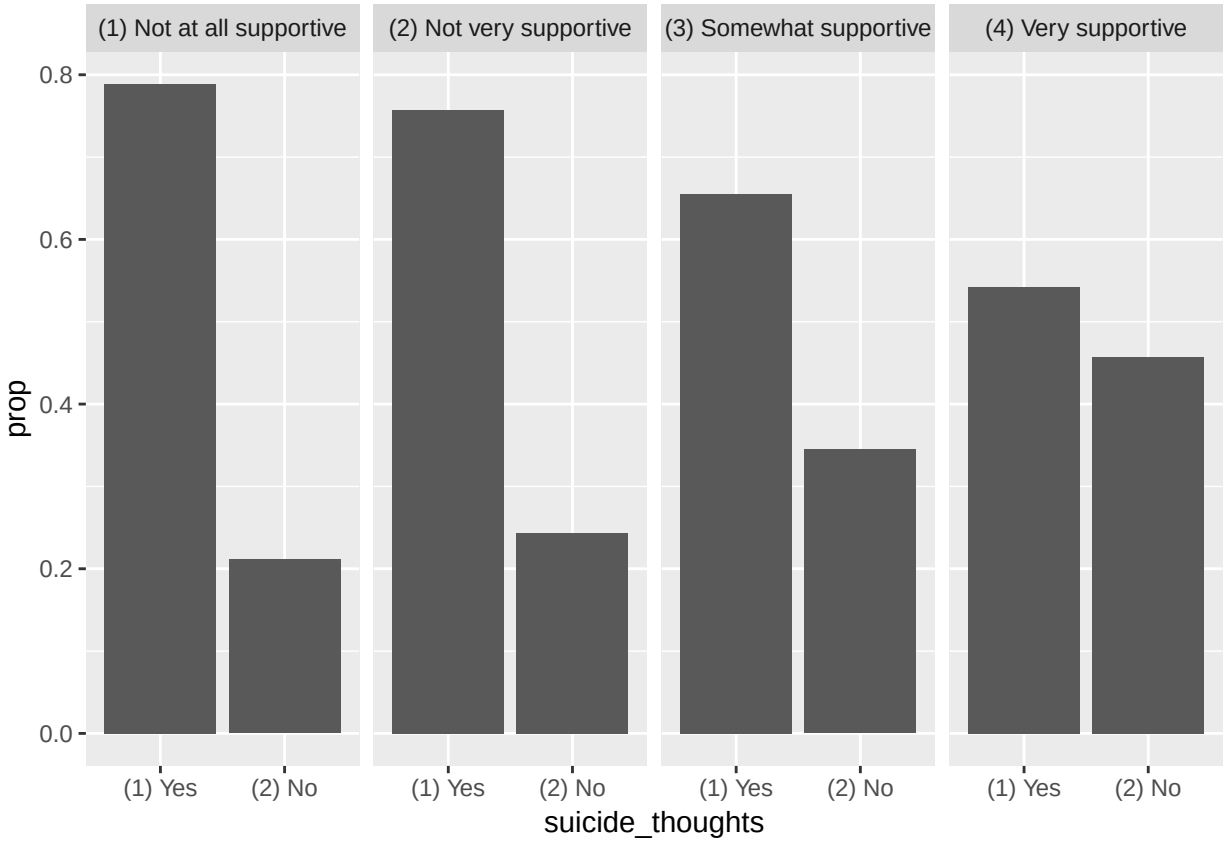
```
## # A tibble: 1 x 1
##   suicide
##     <dbl>
## 1   0.254
```

Comments

25.4% of the trans participants in this survey sample have attempted suicide, which is lower than the national average for trans people.

We would like to know if having a birth family who is supportive of one's gender identity and expression reduces the risk of suicide.

```
this %>%
  filter(!is.na(suicide_thoughts),!is.na(fam_support),
         fam_support != "(5) Not applicable to me") %>%
  ggplot() +
  geom_bar(aes(x=suicide_thoughts,y=..prop..,group=1)) +facet_grid(~fam_support)
```



Comments

We notice that a smaller proportion of participants with supportive families have had suicidal thoughts,indicating a trend between increased familial support and reduced suicidal thoughts. This suggests that having a supportive family can greatly decrease the risk of suicide among trans people.