

hw4_Sayan_Biswas

Sayan Biswas

27 February 2019

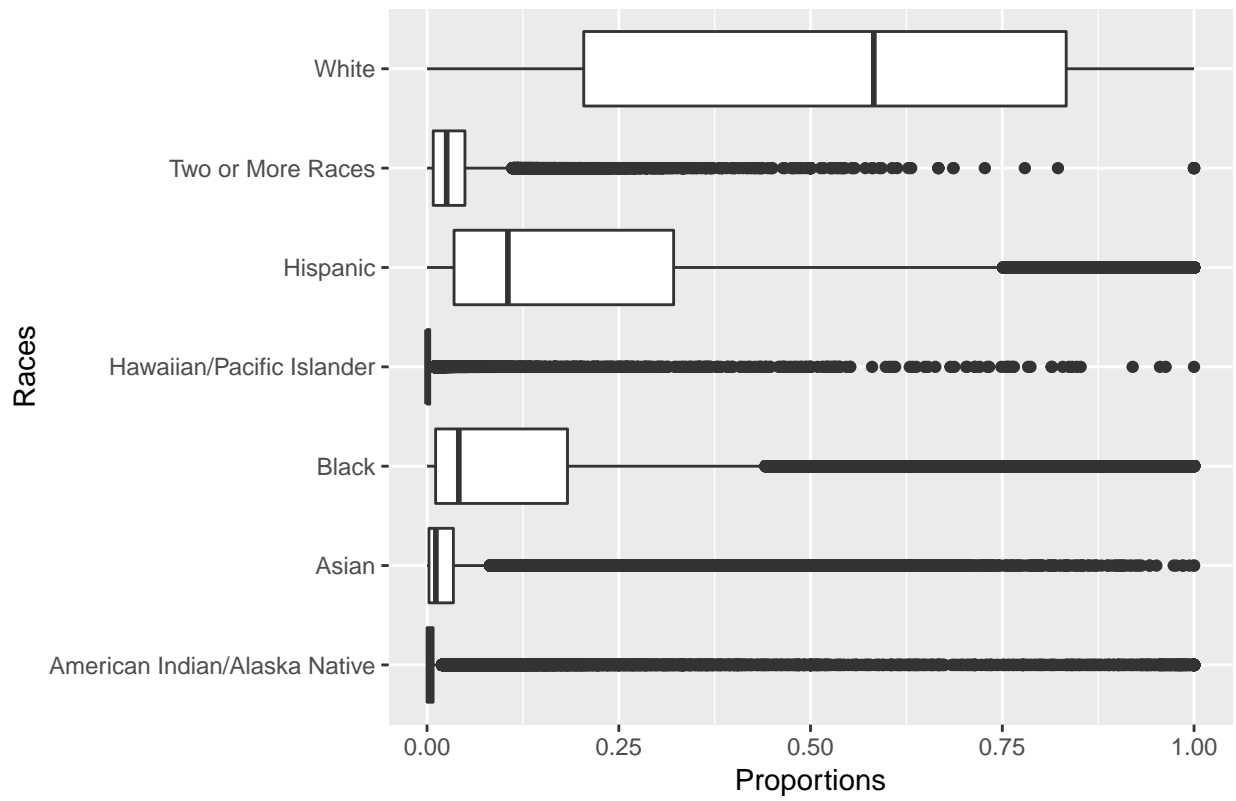
Part A

Problem 1

```
school<- read_csv("CRDC 2015-16 School Data.csv",  
                  na=c("-2", "-5", "-6", "-7", "-8", "-9"),  
                  guess=15000)
```

```
school %>%  
  group_by(LEAID, SCHID)%>%  
  transmute(  
    enr_tot = sum(TOT_ENR_M, TOT_ENR_F, na.rm = TRUE),  
    enr_bl = sum(SCH_ENR_BL_M, SCH_ENR_BL_F, na.rm = TRUE),  
    enr_hisp = sum(SCH_ENR_HI_M, SCH_ENR_HI_F, na.rm = TRUE),  
    enr_am = sum(SCH_ENR_AM_M, SCH_ENR_AM_F, na.rm=TRUE),  
    enr_as = sum(SCH_ENR_AS_M, SCH_ENR_AS_F, na.rm = TRUE),  
    enr_hp = sum(SCH_ENR_HP_M, SCH_ENR_HP_F, na.rm = TRUE),  
    enr_wh = sum(SCH_ENR_WH_M, SCH_ENR_WH_F, na.rm = TRUE),  
    enr_two = sum(SCH_ENR_TR_M, SCH_ENR_TR_F, na.rm = TRUE),  
    Black = enr_bl/enr_tot,  
    Hispanic = enr_hisp/enr_tot,  
    `American Indian/Alaska Native` = enr_am/enr_tot,  
    Asian = enr_as/enr_tot,  
    `Hawaiian/Pacific Islander` = enr_hp/enr_tot,  
    White = enr_wh/enr_tot,  
    `Two or More Races` = enr_two/enr_tot )%>%  
  gather(Black, Hispanic, `American Indian/Alaska Native`, Asian,  
         `Hawaiian/Pacific Islander`, White, `Two or More Races`,  
         key = "Races", value = "Proportions")%>%  
  distinct()%>%  
  ggplot(aes(x=Races, y=Proportions))+  
  geom_boxplot()+  
  labs(title = "Proportions of enrolled students of each race")+  
  coord_flip()
```

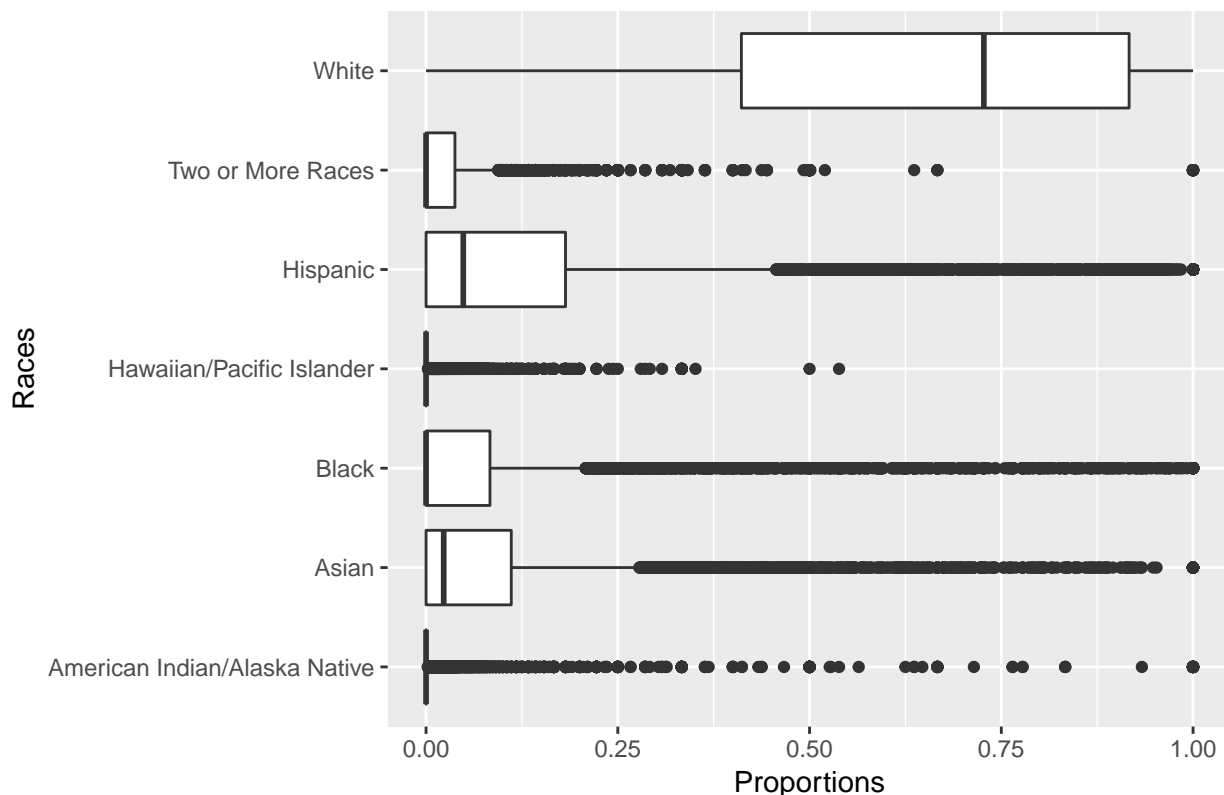
Proportions of enrolled students of each race



Problem 2

```
school%>%
  filter(!is.na(SCH_MATHCLASSES_CALC))%>%
  group_by(LEAID,SCHID)%>%
  transmute(
    cal_tot = sum(TOT_MATHENR_CALC_M,TOT_MATHENR_CALC_F, na.rm = TRUE),
    cal_bl = sum(SCH_MATHENR_CALC_BL_M,SCH_MATHENR_CALC_BL_F, na.rm = TRUE),
    cal_hisp = sum(SCH_MATHENR_CALC_HI_M,SCH_MATHENR_CALC_HI_F,na.rm = TRUE),
    cal_am = sum(SCH_MATHENR_CALC_AM_M,SCH_MATHENR_CALC_AM_F,na.rm = TRUE),
    cal_as = sum(SCH_MATHENR_CALC_AS_M,SCH_MATHENR_CALC_AS_F,na.rm = TRUE),
    cal_hp = sum(SCH_MATHENR_CALC_HP_M,SCH_MATHENR_CALC_HP_F,na.rm = TRUE),
    cal_wh = sum(SCH_MATHENR_CALC_WH_M,SCH_MATHENR_CALC_WH_F,na.rm = TRUE),
    cal_two = sum(SCH_MATHENR_CALC_TR_M,SCH_MATHENR_CALC_TR_F,na.rm = TRUE),
    Black = cal_bl/cal_tot,
    Hispanic = cal_hisp/cal_tot,
    `American Indian/Alaska Native` = cal_am/cal_tot,
    Asian = cal_as/cal_tot,
    `Hawaiian/Pacific Islander` = cal_hp/cal_tot,
    White = cal_wh/cal_tot,
    `Two or More Races`=cal_two/cal_tot)%>%
  gather(Black,Hispanic,`American Indian/Alaska Native`,Asian,
    `Hawaiian/Pacific Islander`,White,`Two or More Races`,
    key = "Races", value = "Proportions")%>%
  distinct()%>%
  ggplot(aes(x=Races,y=Proportions))+
  geom_boxplot()+
  coord_flip()+
  labs(title = "Prop. of students of each race enrolled in Calculus class")
```

Prop. of students of each race enrolled in Calculus class



1. The proportion of White students enrolled in each school is greater than any of the other races. The median proportion of white students enrolled in calculus class is greater than the median proportion of students of all other races.
2. The median proportion of students of each race enrolled in each school in descending order: White students > Hispanic students > Black Students > Asian students > American Indian/Alaska Native > Hawaiian/Pacific Islander.
3. The median proportion of students of each race enrolled in calculus class in each school in descending order: White students > Hispanic students > Asian students > Black Students > American Indian/Alaska Native students > Hawaiian/Pacific Islander students.
4. Though the median proportion of Black students enrolled in each school is greater than the median proportion of Asian students enrolled, the median proportion of Black students is less than the median proportion of Asian students enrolled in Calculus class.
5. The median proportion of two or more races students enrolled in Calculus class is close to zero and less than the median proportion of Asian students, though the median proportion of the two or more race students enrolled in each school is not zero and slightly greater than the proportion of Asian students.
6. The median proportion of American Indian and Hawaiian/Pacific Islander students enrolled in each school is very less and also the median proportion of these students enrolled for each Calculus class in each school is very less close to zero.

Part B

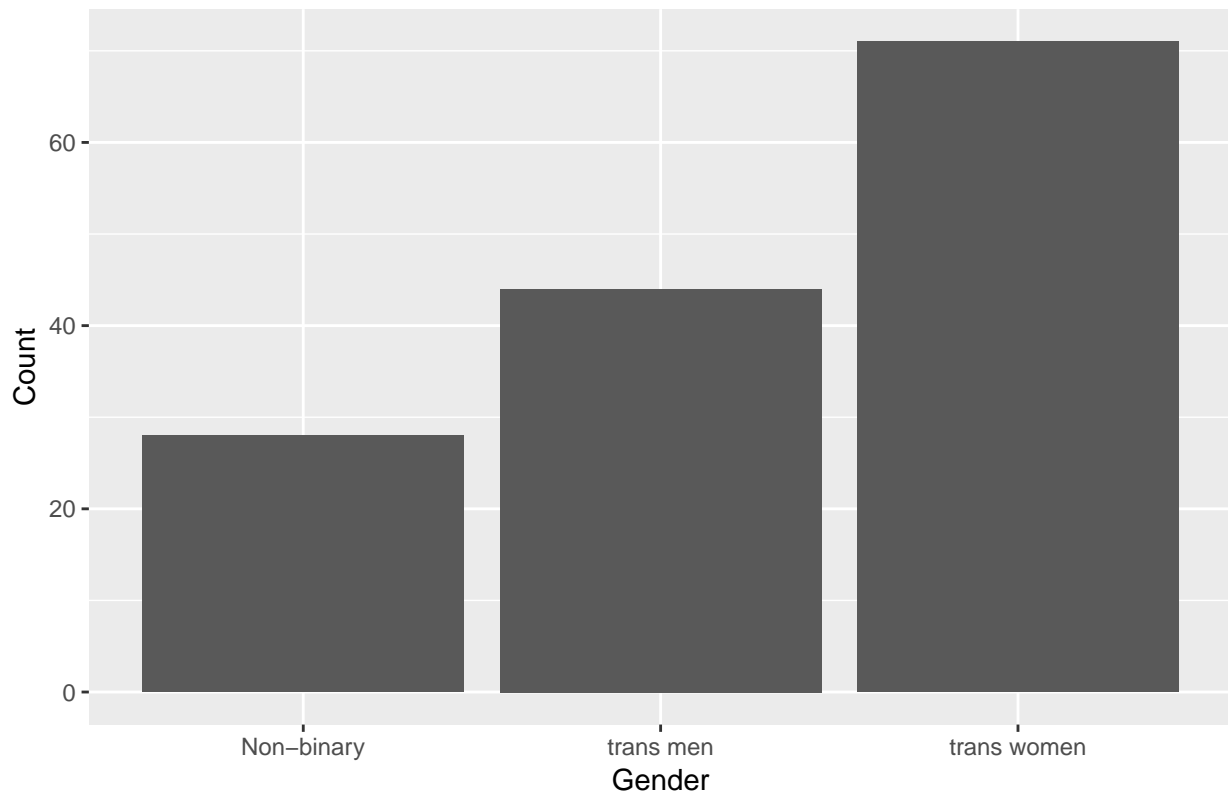
Problem 3

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

```
`da31721.0001`%>%
select(Q5,Q6)%>%
separate(Q5,into = c("code1","birth_gender"), sep = " ", extra = "merge")%>%
separate(Q6, into = c("code2","present_gender"), sep = " ", extra = "merge")%>%
select(-code1,-code2)%>%
mutate(gender = ifelse(birth_gender=="Male" & present_gender=="Woman",
                      "trans women",
                      ifelse(birth_gender=="Female" & present_gender=="Man",
                              "trans men",
                              ifelse(present_gender %in% c('Gender Queer','Androgynous'),
                                      "Non-binary","Unknown"))))%>%

filter(gender!="Unknown")%>%
count(gender)%>%
ggplot(aes(x=gender,y=n))+
geom_col()+
labs(title = "Participants of each gender",
      x = "Gender",
      y = "Count")
```

Participants of each gender

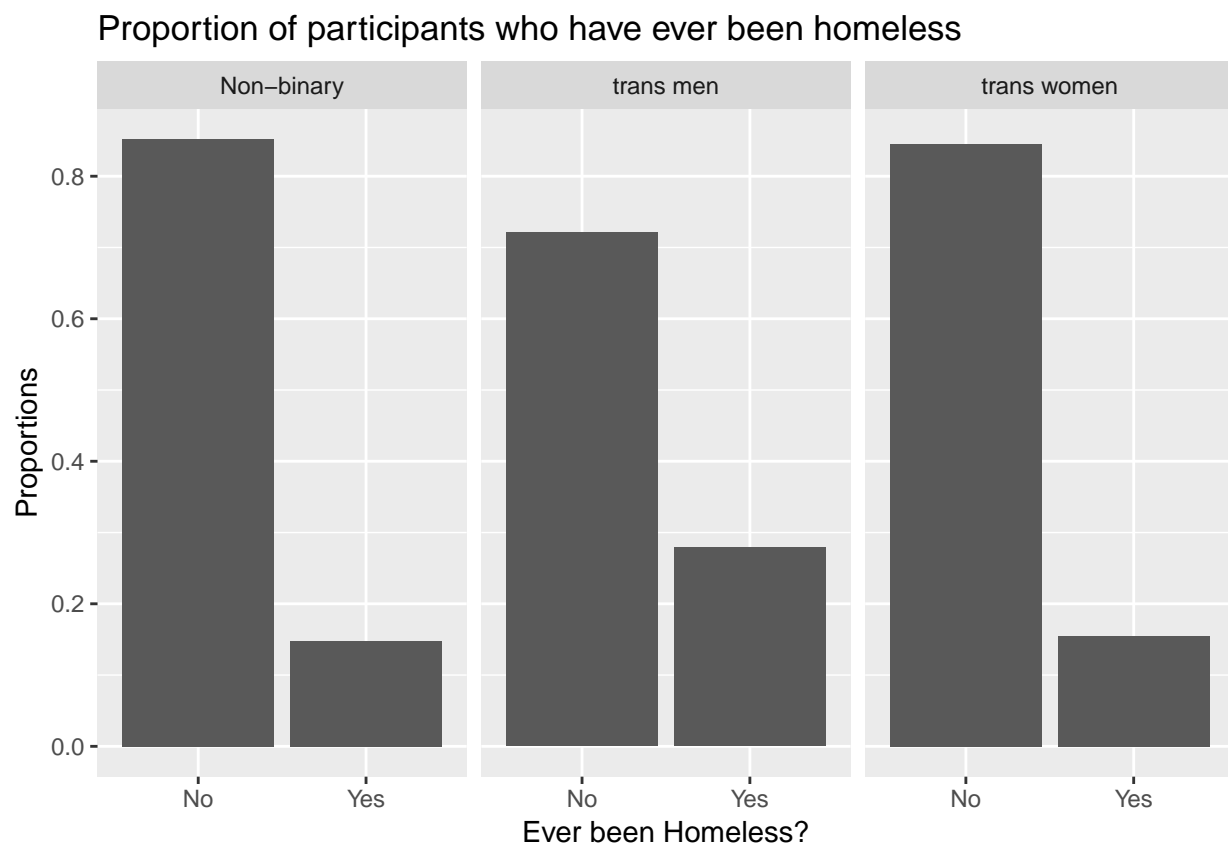


```

`da31721.0001`%>%
  separate(Q5,into = c("code1","birth_gender"), sep = " ", extra = "merge")%>%
  separate(Q6, into = c("code2","present_gender"), sep = " ", extra = "merge")%>%
  select(-code1,-code2)%>%
  mutate(gender = ifelse(birth_gender=="Male" & present_gender=="Woman",
                        "trans women",
                        ifelse(birth_gender=="Female" & present_gender=="Man",
                              "trans men",
                              ifelse(present_gender %in% c('Gender Queer','Androgynous'),
                                    "Non-binary","Unknown"))))%>%

  filter(gender!="Unknown")%>%
  rename("Homeless"= Q88)%>%
  separate(Homeless,into = c("code3","value"), sep = " ")%>%
  mutate(value = ifelse(is.na(value),"Unknown",value))%>%
  filter(value!="Unknown")%>%
  ggplot(aes(x=value,y=..prop..,group=1))+
  geom_bar()+
  facet_wrap(~gender)+
  labs(title="Proportion of participants who have ever been homeless",
       x="Ever been Homeless?",
       y="Proportions")

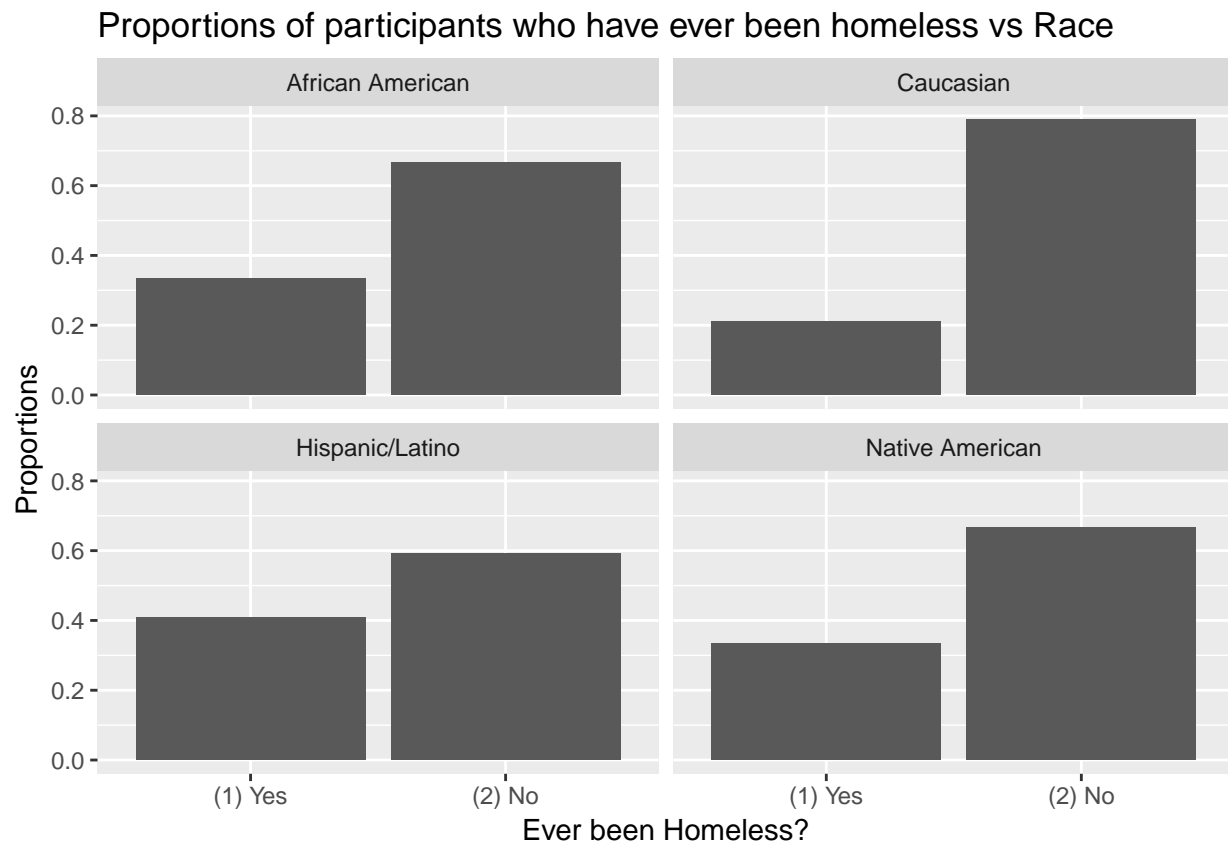
```



The proportion of Non-binary and trans women participants ever been homeless is greater than twice the proportion of general U.S. population who have ever been homeless. The proportion of Trans men is far more worse as it is greater than four times the proportion of general U.S. population who have ever been homeless. The proportion of trans men participants ever been homeless is greater than the proportion of trans women participants ever been homeless which is greater than the proportion of non-binary participants.

Problem 4

```
`da31721.0001`%>%
  mutate(`African American`=ifelse(D9_1=="(1) Selected",1,0))%>%
  mutate(Caucasian=ifelse(D9_2=="(1) Selected",1,0))%>%
  mutate(`Hispanic/Latino`=ifelse(D9_3=="(1) Selected",1,0))%>%
  mutate(`Native American`=ifelse(D9_4=="(1) Selected",1,0))%>%
  gather(`African American`,Caucasian,`Hispanic/Latino`,
        `Native American`,key="Races",value="Selected")%>%
  filter(Selected==1)%>%
  filter(!is.na(Q88))%>%
  ggplot(aes(x=Q88,y=..prop..,group=1))+
  geom_bar()+
  facet_wrap(~Races)+
  labs(title = "Proportions of participants who have ever been homeless vs Race",
       x="Ever been Homeless?",
       y="Proportions")
```



1. Overall percentage of homelessness among these four races is higher compared to the general US population.
2. The proportion of Hispanic/Latino participants ever been homeless is higher than the proportion of Native American and African American which are higher than Caucasian participants.
3. The proportion of Hispanic/Latino participants is 40% which is nearly 6 times the 6.2% of the general U.S. population has ever been homeless.
4. The proportion of Native American participants and African American participants is nearly 5 times 6.2% of the general U.S. population has ever been homeless.
5. The proportion of Caucasian participants is nearly 3 times 6.2% of the general U.S. population has ever been homeless.

Problem 5

```
`da31721.0001`%>%  
  summarise("Total proportion of survey participants who attempted suicide"  
    =sum(Q133=="(1) Yes",na.rm = TRUE)/n())
```

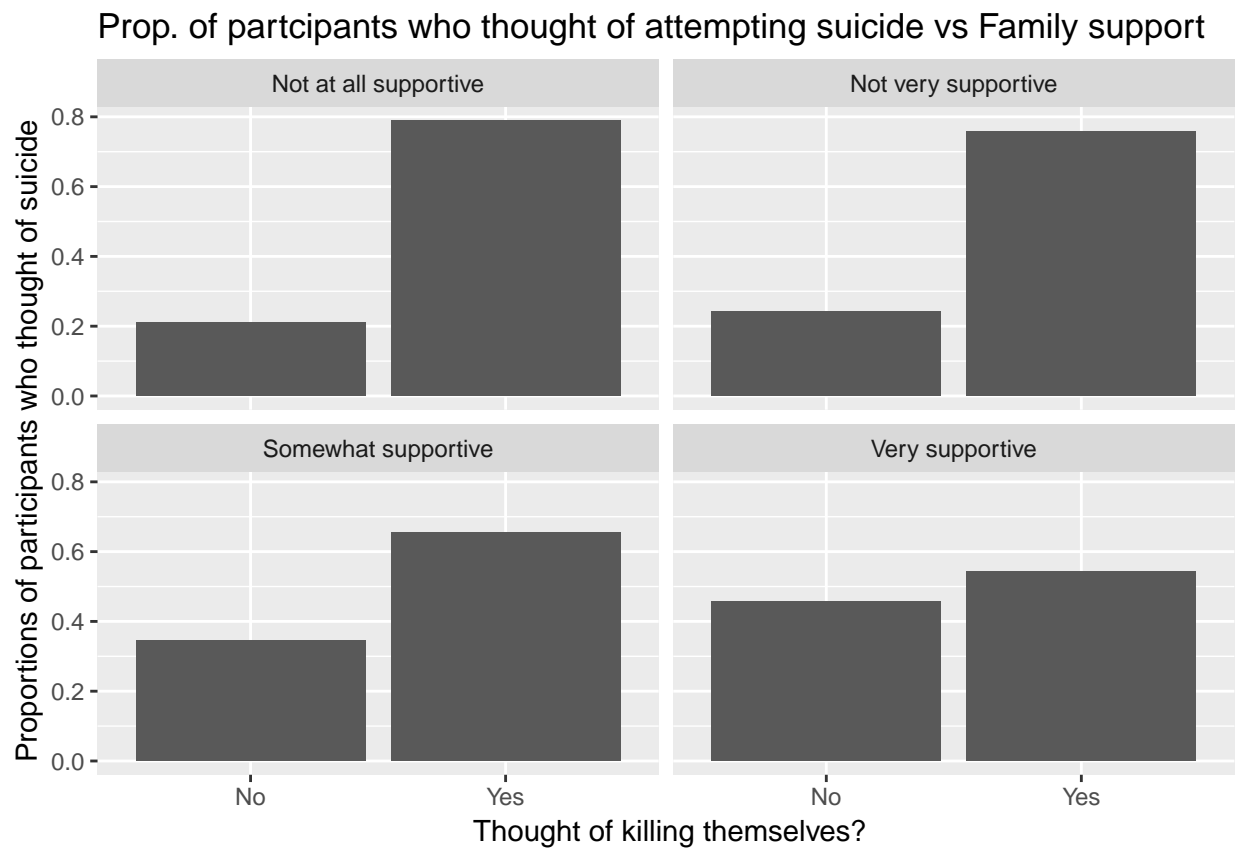
```
## Total proportion of survey participants who attempted suicide  
## 1 0.2542857
```

The total proportion of survey participants who attempted suicide is ~ 0.254 or 25.4% which is less than the national average for trans people which is 40%. But the total proportion of survey participants who attempted suicide is far more worse when compared to the national average, which is nearly five times the attempted suicide rate of the general U.S. population (4.6%).

```

`da31721.0001`%>%
  separate(Q131,into = c("code131","value131"), sep = " ")%>%
  filter(!is.na(value131))%>%
  separate(Q119,into = c("code119","value119"), sep = " ",extra = "merge")%>%
  filter(value119!="Not applicable to me",!(is.na(value119)))%>%
  ggplot(aes(x=value131,y=..prop..,group=1))+
  geom_bar()+
  facet_wrap(~value119)+
  labs(title =
    "Prop. of participants who thought of attempting suicide vs Family support",
    x = "Thought of killing themselves?",
    y = "Proportions of participants who thought of suicide")

```



Clearly family support plays a vital role in reducing the proportion of participants who thought of attempting suicide as the proportion of participants who thought of attempting a suicide with a very supportive family is lower compared to the proportion of participants who thought of attempting a suicide with having a not at all supportive family, a not very supportive supportive family and also lower compared to a somewhat supportive family.