

2020MCS120003_LabAssignment4

Amith C A

04/10/2020

Step 1: Changing characters to factors

```
flying_etiquette<-read.csv("flying_etiquette.csv",na.strings=c(""),stringsAsFactors=TRUE)
df_fe<-filter(flying_etiquette,How.often.do.you.travel.by.plane.!="Never")
glimpse(df_fe)
```

```
## Rows: 874
## Columns: 27
## $ RespondentID
## $ How.often.do.you.travel.by.plane.
## $ Do.you.ever.recline.your.seat.when.you.fly.
## $ How.tall.are.you.
## $ Do.you.have.any.children.under.18.
## $ In.a.row.of.three.seats..who.should.get.to.use.the.two.arm.rests.
## $ In.a.row.of.two.seats..who.should.get.to.use.the.middle.arm.rest.
## $ Who.should.have.control.over.the.window.shade.
## $ Is.it.rude.to.move.to.an.unsold.seat.on.a.plane.
## $ Generally.speaking..is.it.rude.to.say.more.than.a.few.words.to.the.stranger.sitting.next.to.you.on.
## $ On.a.6.hour.flight.from.NYC.to.LA..how.many.times.is.it.acceptable.to.get.up.if.you.re.not.in.an.a
## $ Under.normal.circumstances..does.a.person.who.reclines.their.seat.during.a.flight.have.any.obligat
## $ Is.it.rude.to.recline.your.seat.on.a.plane.
## $ Given.the.opportunity..would.you.eliminate.the.possibility.of.reclining.seats.on.planes.entirely.
## $ Is.it.rude.to.ask.someone.to.switch.seats.with.you.in.order.to.be.closer.to.friends.
## $ Is.it.rude.to.ask.someone.to.switch.seats.with.you.in.order.to.be.closer.to.family.
## $ Is.it.rude.to.wake.a.passenger.up.if.you.are.trying.to.go.to.the.bathroom.
## $ Is.it.rude.to.wake.a.passenger.up.if.you.are.trying.to.walk.around.
## $ In.general..is.it.rude.to.bring.a.baby.on.a.plane.
## $ In.general..is.it.rude.to.knowingly.bring.unruly.children.on.a.plane.
## $ Have.you.ever.used.personal.electronics.during.take.off.or.landing.in.violation.of.a.flight.attend
## $ Have.you.ever.smoked.a.cigarette.in.an.airplane.bathroom.when.it.was.against.the.rules.
## $ Gender
## $ Age
## $ Household.Income
## $ Education
## $ Location..Census.Region.
```

Step 2 : Tidying data

```
col_names_rude<-names(df_fe[,str_detect(names(df_fe),"rude")])
df_fe_long<-gather(df_fe,"response_var","values",c(2:(ncol(df_fe)-5)))
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
head(df_fe_long)
```

```
##   RespondentID Gender   Age Household.Income      Education
## 1   3436139758  <NA>  <NA>             <NA>             <NA>
## 2   3434278696  Male 30-44             <NA> Graduate degree
## 3   3434275578  Male 30-44 $100,000 - $149,999 Bachelor degree
## 4   3434268208  Male 30-44    $0 - $24,999 Bachelor degree
## 5   3434250245  Male 30-44  $50,000 - $99,999 Bachelor degree
## 6   3434245875  Male 30-44  $25,000 - $49,999 Graduate degree
##   Location..Census.Region.      response_var
## 1             <NA> How.often.do.you.travel.by.plane.
## 2             Pacific How.often.do.you.travel.by.plane.
## 3             Pacific How.often.do.you.travel.by.plane.
## 4             Pacific How.often.do.you.travel.by.plane.
## 5             Pacific How.often.do.you.travel.by.plane.
## 6   East North Central How.often.do.you.travel.by.plane.
##               values
## 1  Once a year or less
## 2  Once a year or less
## 3  Once a year or less
## 4  Once a year or less
## 5  Once a month or less
## 6  Once a year or less
```

Step 3 : Cleaning up strings

```
df_fe_long$response_var[str_detect(df_fe_long$response_var,"rude.to.")]<-str_remove(df_fe_long$response_var,"rude.to.")
df_fe_long$response_var[str_detect(df_fe_long$response_var,".on.a.plane")]<-str_remove(df_fe_long$response_var,".on.a.plane")
unique(df_fe_long$response_var)
```

```
## [1] "How.often.do.you.travel.by.plane."
## [2] "Do.you.ever.recline.your.seat.when.you.fly."
## [3] "How.tall.are.you."
## [4] "Do.you.have.any.children.under.18."
## [5] "In.a.row.of.three.seats..who.should.get.to.use.the.two.arm.rests."
## [6] "In.a.row.of.two.seats..who.should.get.to.use.the.middle.arm.rest."
## [7] "Who.should.have.control.over.the.window.shade."
## [8] "move.to.an.unsold.seat"
## [9] "say.more.than.a.few.words.tothe.stranger.sitting.next.to.you"
## [10] "On.a.6.hour.flight.from.NYC.to.LA..how.many.times.is.it.acceptable.to.get.up.if.you.re.not.in.the.airplane."
## [11] "Under.normal.circumstances..does.a.person.who.reclines.their.seat.during.a.flight.have.any.obligations?"
## [12] "recline.your.seat"
## [13] "Given.the.opportunity..would.you.eliminate.the.possibility.of.reclining.seats.on.planes.entirely?"
```

```
## [14] "ask.someone.to.switch.seats.with.you.in.order.to.be.closer.to.friends."
## [15] "ask.someone.to.switch.seats.with.you.in.order.to.be.closer.to.family."
## [16] "wake.a.passenger.up.if.you.are.trying.to.go.to.the.bathroom."
## [17] "wake.a.passenger.up.if.you.are.trying.to.walk.around."
## [18] "bring.a.baby"
## [19] "knowingly.bring.unruly.children"
## [20] "Have.you.ever.used.personal.electronics.during.take.off.or.landing.in.violation.of.a.flight.at"
## [21] "Have.you.ever.smoked.a.cigarette.in.an.airplane.bathroom.when.it.was.against.the.rules."
```

Step 4 : Dichotomizing variables

```
dim(df_fe_long[is.na(df_fe_long$values),])
```

```
## [1] 422 8
```

```
df_fe_long_NA<-df_fe_long[is.na(df_fe_long$values),]
df_fe_long<-df_fe_long[!is.na(df_fe_long$values),]
dim(df_fe_long[is.na(df_fe_long$values),])
```

```
## [1] 0 8
```

```
dim(df_fe_long_NA)
```

```
## [1] 422 8
```

```
df_fe_long$rude<-0
df_fe_long[str_detect(df_fe_long$values,c("No, not at all rude","Not, at all rude")),]$rude<-1
df_fe_long %>% group_by(df_fe_long$rude) %>% summarise(n=n())
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
## # A tibble: 2 x 2
##   'df_fe_long$rude'      n
##           <dbl> <int>
## 1             0 16161
## 2             1  1771
```

```
head(arrange(df_fe_long,desc(rude)))
```

```
##   RespondentID Gender  Age  Household.Income
## 1  3434278696   Male 30-44          <NA>
## 2  3434268208   Male 30-44      $0 - $24,999
## 3  3434245875   Male 30-44  $25,000 - $49,999
## 4  3434218031   Male 30-44      $0 - $24,999
## 5  3434122066   Male 30-44 $100,000 - $149,999
## 6  3434107997   Male 18-29  $25,000 - $49,999
##                                     Education Location..Census.Region.
## 1                               Graduate degree                               Pacific
```

```
## 2 Bachelor degree Pacific
## 3 Graduate degree East North Central
## 4 Bachelor degree New England
## 5 Some college or Associate degree Middle Atlantic
## 6 Bachelor degree South Atlantic
## response_var
## 1 say.more.than.a.few.words.to.the.stranger.sitting.next.to.you
## 2 say.more.than.a.few.words.to.the.stranger.sitting.next.to.you
## 3 say.more.than.a.few.words.to.the.stranger.sitting.next.to.you
## 4 say.more.than.a.few.words.to.the.stranger.sitting.next.to.you
## 5 say.more.than.a.few.words.to.the.stranger.sitting.next.to.you
## 6 say.more.than.a.few.words.to.the.stranger.sitting.next.to.you
## values rude
## 1 No, not at all rude 1
## 2 No, not at all rude 1
## 3 No, not at all rude 1
## 4 No, not at all rude 1
## 5 No, not at all rude 1
## 6 No, not at all rude 1
```

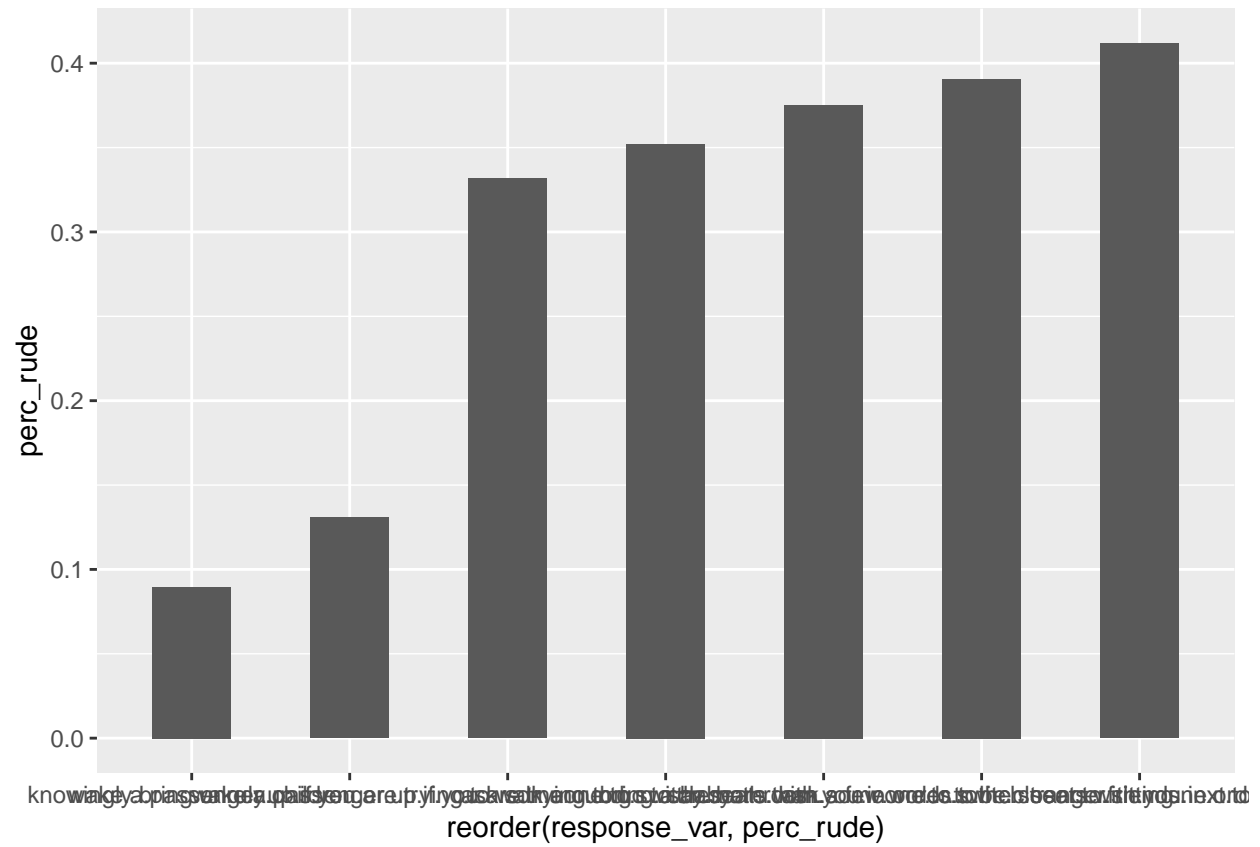
Step 5 : Summarizing data

```
rude_behaviors<-df_fe_long %>%group_by(response_var) %>% summarise(perc_rude=mean(rude), .groups = 'drop')
rude_behaviors
```

```
## # A tibble: 21 x 2
## response_var perc_rude
## <chr> <dbl>
## 1 ask.someone.to.switch.seats.with.you.in.order.to.be.closer.to.fami~ 0.412
## 2 ask.someone.to.switch.seats.with.you.in.order.to.be.closer.to.frie~ 0.375
## 3 bring.a.baby 0.352
## 4 Do.you.ever.recline.your.seat.when.you.fly. 0
## 5 Do.you.have.any.children.under.18. 0
## 6 Given.the.opportunity..would.you.eliminate.the.possibility.of.recl~ 0
## 7 Have.you.ever.smoked.a.cigarette.in.an.airplane.bathroom.when.it.w~ 0
## 8 Have.you.ever.used.personal.electronics.during.take.off.or.landing~ 0
## 9 How.often.do.you.travel.by.plane. 0
## 10 How.tall.are.you. 0
## # ... with 11 more rows
```

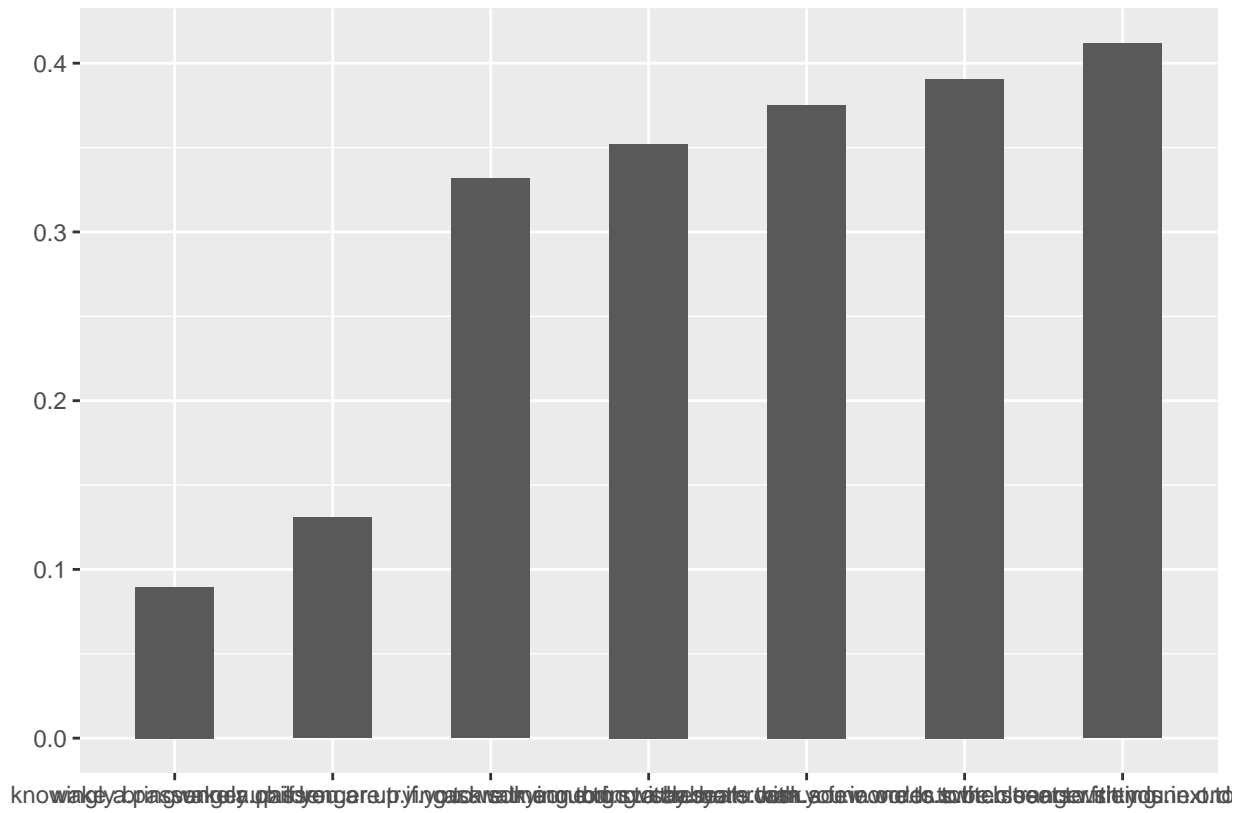
Step 6 : Create a plot

```
initial_plot<-ggplot(filter(rude_behaviors,perc_rude>0),aes(x=reorder(response_var,perc_rude),y=perc_rude))
initial_plot
```



Step 7 : Fixing Labels

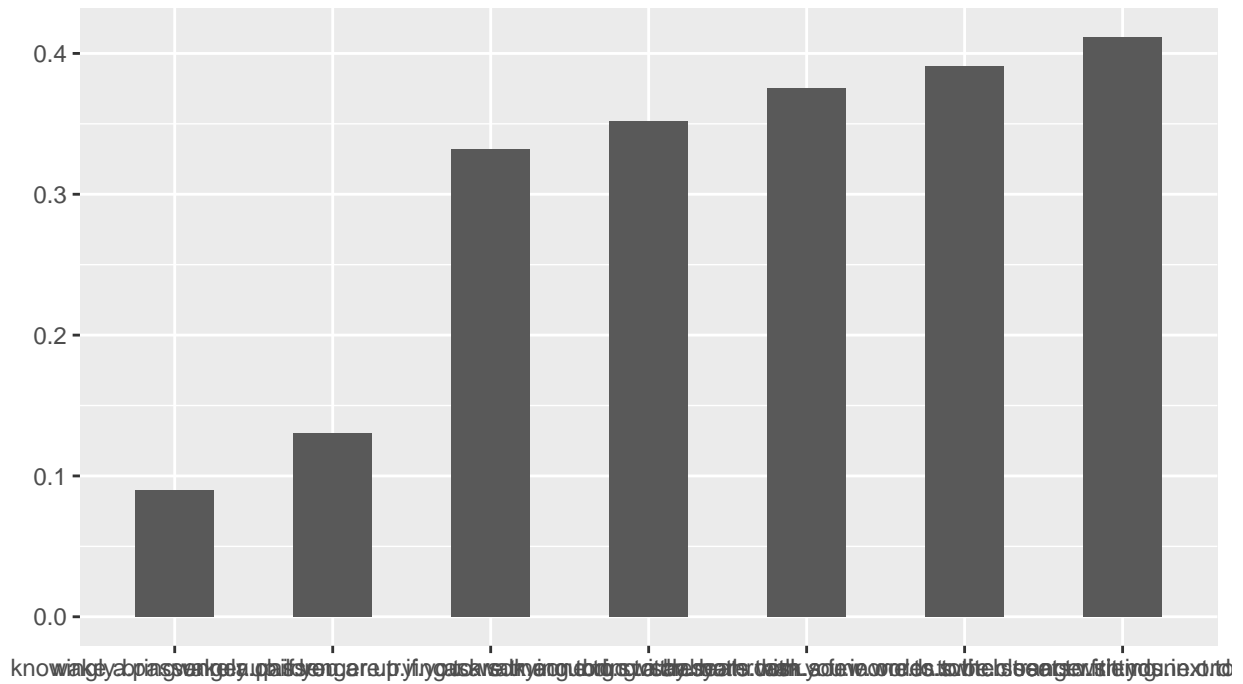
```
new_plot<-initial_plot+labs(x="",y="")
new_plot
```



```
titled_plot<-initial_plot+labs(title="Hell Is Other People In A Pressurized Metal Tube",subtitle="Percentage of Passengers Who Survived or Died in the Titanic Disaster")
titled_plot
```

Hell Is Other People In A Pressurized Metal Tube

Percentage of 874 air-passenger respondents who said action is very or somewhat rude



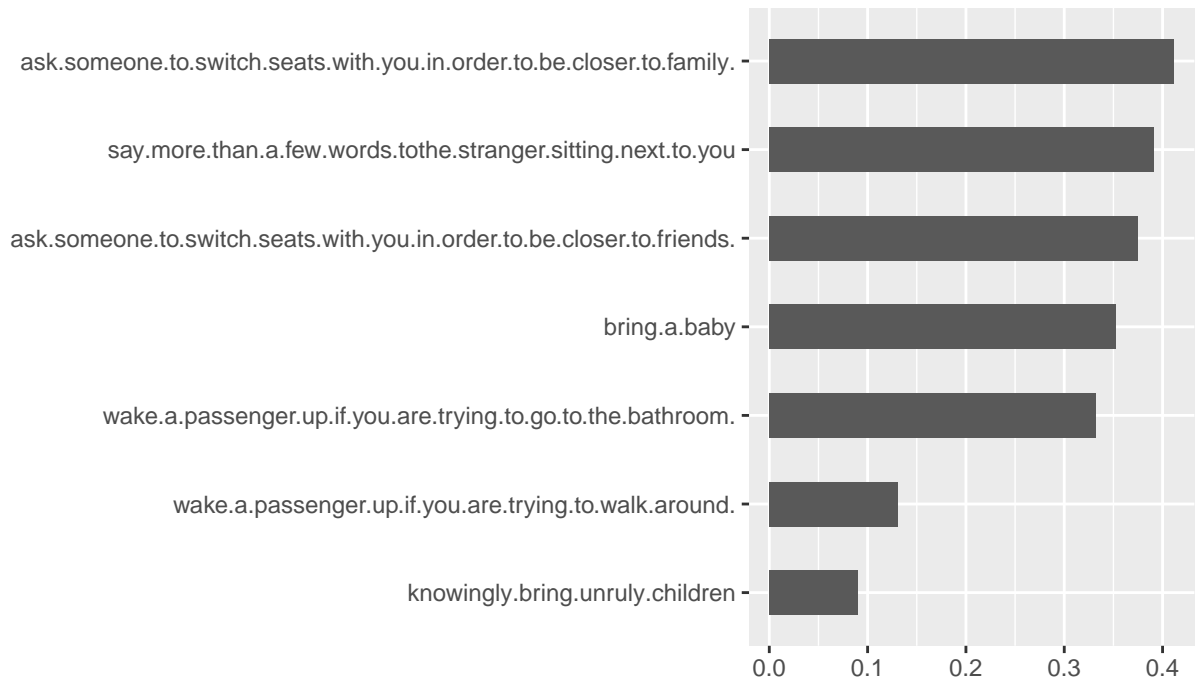
Source: Survey Monkey Audience

Step 8 : Flipping coordinates

```
titled_plot+coord_flip()
```

Hell Is Other People In A Pres

Percentage of 874 air-passenger re

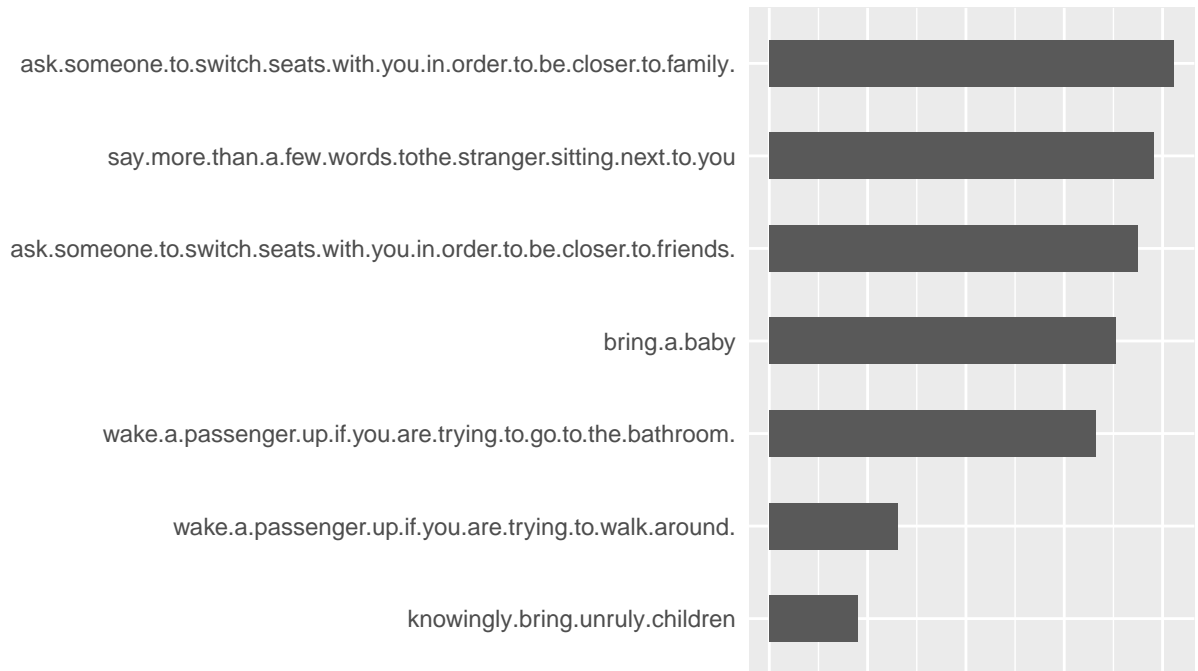


Source: Survey Monkey Audience

```
titled_plot+coord_flip()+ theme(  
  axis.text.x = element_blank(),  
  axis.ticks = element_blank())
```


Hell Is Other People In A Pres

Percentage of 874 air-passenger re



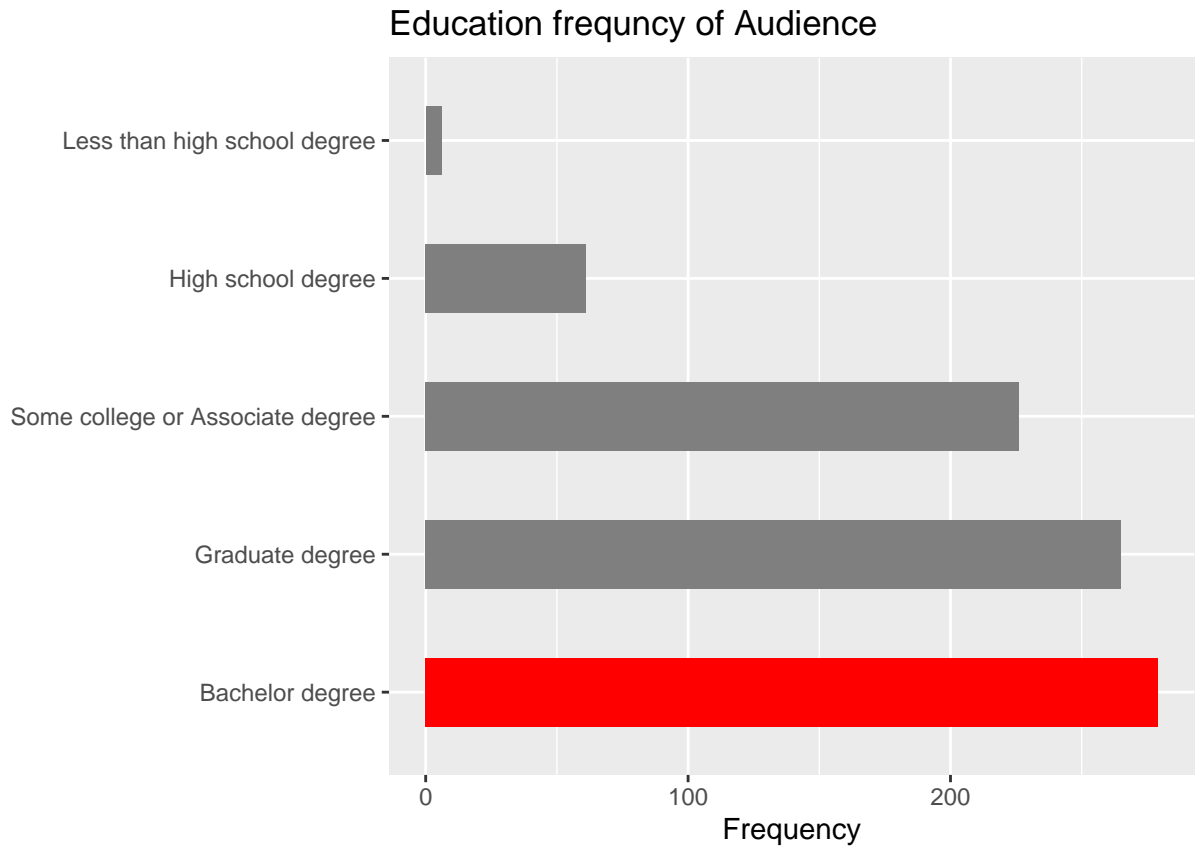
Source: Survey Monkey Audience

Step 9 : Find out

```
df_edu<-df_fe%>%filter(!is.na(Education)) %>% group_by(Education) %>% summarize(freq=n())
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
ggplot(df_edu,aes(x=freq,y=reorder(Education,-freq), fill=factor(ifelse(Education=="Bachelor degree","H
```

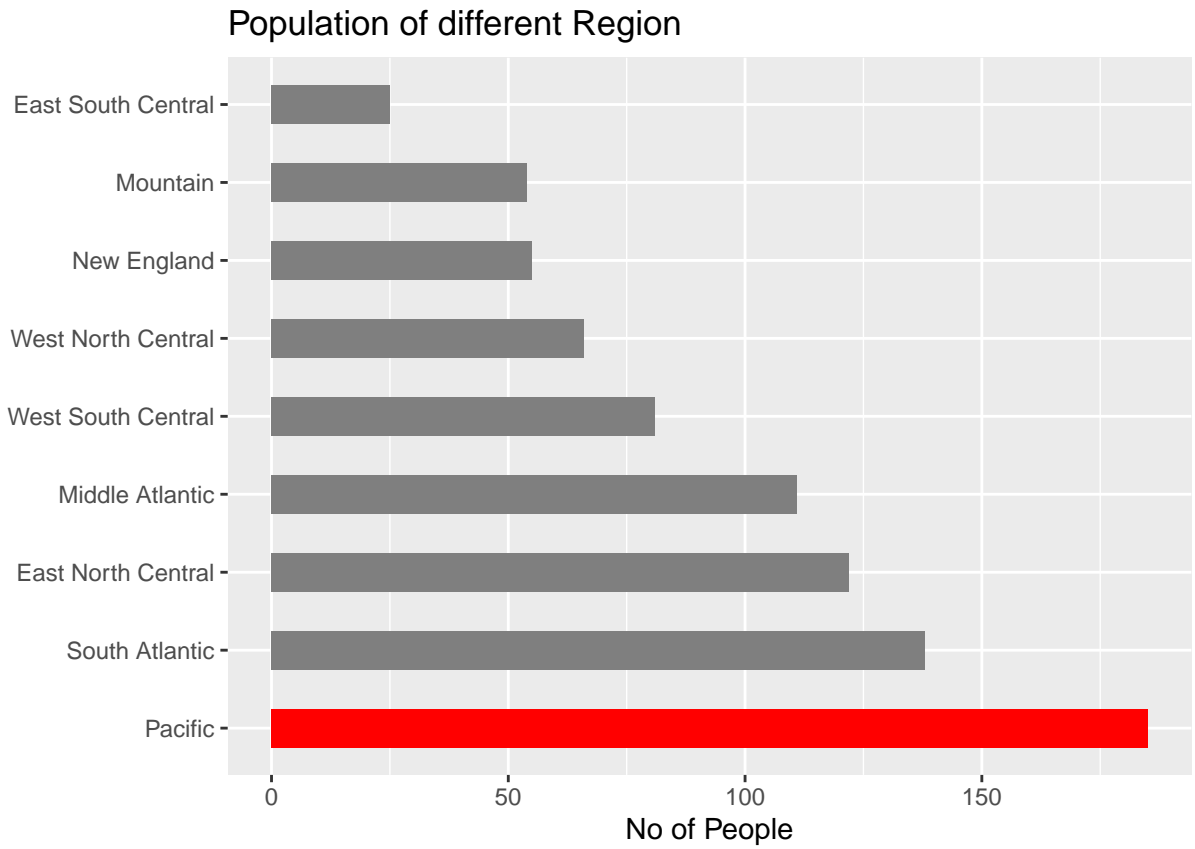


Highest education frequency is for **Bachelors degree**

```
df_loc<-df_fe%>%filter(!is.na(Location..Census.Region.)) %>% group_by(Location..Census.Region.) %>% sum
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
ggplot(df_loc,aes(x=pop,y=reorder(Location..Census.Region.,-pop), fill=factor(ifelse(Location..Census.R
```



Most people are from the **Pacific** region.