Social Inclusion Analysis

Yining Hua

12/10/2021

```
dat <- read_csv("preliminary_data.csv")
```

```
## Rows: 33697 Columns: 24
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (10): gender, education, marriage, migration.scale, job, hangouts, willi...
## dbl (14): participant, ethnicity, expence, income, worked_before5.1, health,...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
col_names <- c("job","marriage","diabete.or.hypertension","group","insuranced","gender","worked_before5
dat[,col_names] <- lapply(dat[,col_names] , factor)

dat$money.left <- dat$income - dat$expence
dat$money.left <- (dat$money.left - mean(dat$money.left, na.rm = TRUE)) / sd(dat$money.left, na.rm = TRU
dat$participant <- as.character(dat$participant)

dat$health <- as.factor(dat$health)
dat$migration.scale <- as.factor(dat$migration.scale)
dat$education.group <- as.factor(dat$education)
dat$health_combined <- as.factor(dat$health_combined)
```

```
head(dat)
```

```
## # A tibble: 6 x 26
##   participant gender ethnicity education marriage migration.scale expence income
##   <chr>       <fct>      <dbl> <chr>     <fct>    <fct>             <dbl>  <dbl>
## 1 0           female         1 highscho~ just ma~ interstate        10000     NA
## 2 2           female         1 highscho~ just ma~ interstate        40000     NA
## 3 4           female         1 junior c~ just ma~ interstate         9000     NA
## 4 16          male           1 highscho~ just ma~ interstate         2000 -10000
## 5 18          female         1 midschool just ma~ intercounty        5000  -8000
## 6 25          male           1 highscho~ just ma~ intercity          6000  -5000
## # ... with 18 more variables: worked_before5.1 <fct>, job <fct>,
## #   hangouts <chr>, willing.to.movein <fct>, willing.to.stay <fct>,
## #   health <fct>, diabete.or.hypertension <fct>, group <fct>,
```

```
## #   participated.in.group.activity <dbl>, like.current.city <dbl>,
## #   natives.like.me <dbl>, natives.lookdown.me <dbl>,
## #   previous.customs.better <dbl>, i.am.native <dbl>, insuranced <fct>,
## #   health_combined <fct>, money.left <dbl>, education.group <fct>
```

```
# dat$natives_inclusion <- dat$natives.like.me-dat$natives.lookdown.me
dat$city_inclusion <- dat$like.current.city-dat$previous.customs.better+dat$i.am.native + dat$natives.li
# dat$tendency.livehere <- dat$willing.to.movein + dat$willing.to.stay
# dat$loneliness.level <- (dat$loneliness.level-min(dat$loneliness.level, na.rm=TRUE))/(max(dat$lonelin

dat
```

```
## # A tibble: 33,697 x 27
##    participant gender ethnicity education      marriage  migration.scale expence
##    <chr>       <fct>      <dbl> <chr>          <fct>     <fct>             <dbl>
##  1 0           female         1 highschool     just mar~ interstate        10000
##  2 2           female         1 highschool     just mar~ interstate        40000
##  3 4           female         1 junior college just mar~ interstate         9000
##  4 16          male           1 highschool     just mar~ interstate         2000
##  5 18          female         1 midschool      just mar~ intercounty        5000
##  6 25          male           1 highschool     just mar~ intercity          6000
##  7 31          female         1 midschool      just mar~ intercity          2500
##  8 41          female         1 midschool      just mar~ interstate         3000
##  9 44          male           1 midschool      remarried interstate         3166
## 10 59          male           1 midschool      just mar~ intercounty        3000
## # ... with 33,687 more rows, and 20 more variables: income <dbl>,
## #   worked_before5.1 <fct>, job <fct>, hangouts <chr>, willing.to.movein <fct>,
## #   willing.to.stay <fct>, health <fct>, diabete.or.hypertension <fct>,
## #   group <fct>, participated.in.group.activity <dbl>, like.current.city <dbl>,
## #   natives.like.me <dbl>, natives.lookdown.me <dbl>,
## #   previous.customs.better <dbl>, i.am.native <dbl>, insuranced <fct>,
## #   health_combined <fct>, money.left <dbl>, education.group <fct>, ...
```

```
## regroup education
dat$education.group <- NA
dat$education.group[dat$education == "no education"] <- "low"
dat$education.group[dat$education == "primary school"] <- "low"
dat$education.group[dat$education == "midschool"] <- "middle"
dat$education.group[dat$education == "highschool"] <- "middle"
dat$education.group[dat$education =="junior college"] <- "middle"
dat$education.group[dat$education == "college"] <- "high"
dat$education.group[dat$education == "grad"] <- "high"
```

```
## regroup ethnicity
#dat$ethnicity.group <- "other"
#dat$ethnicity.group[dat$ethnicity == 1] <- "han"
```

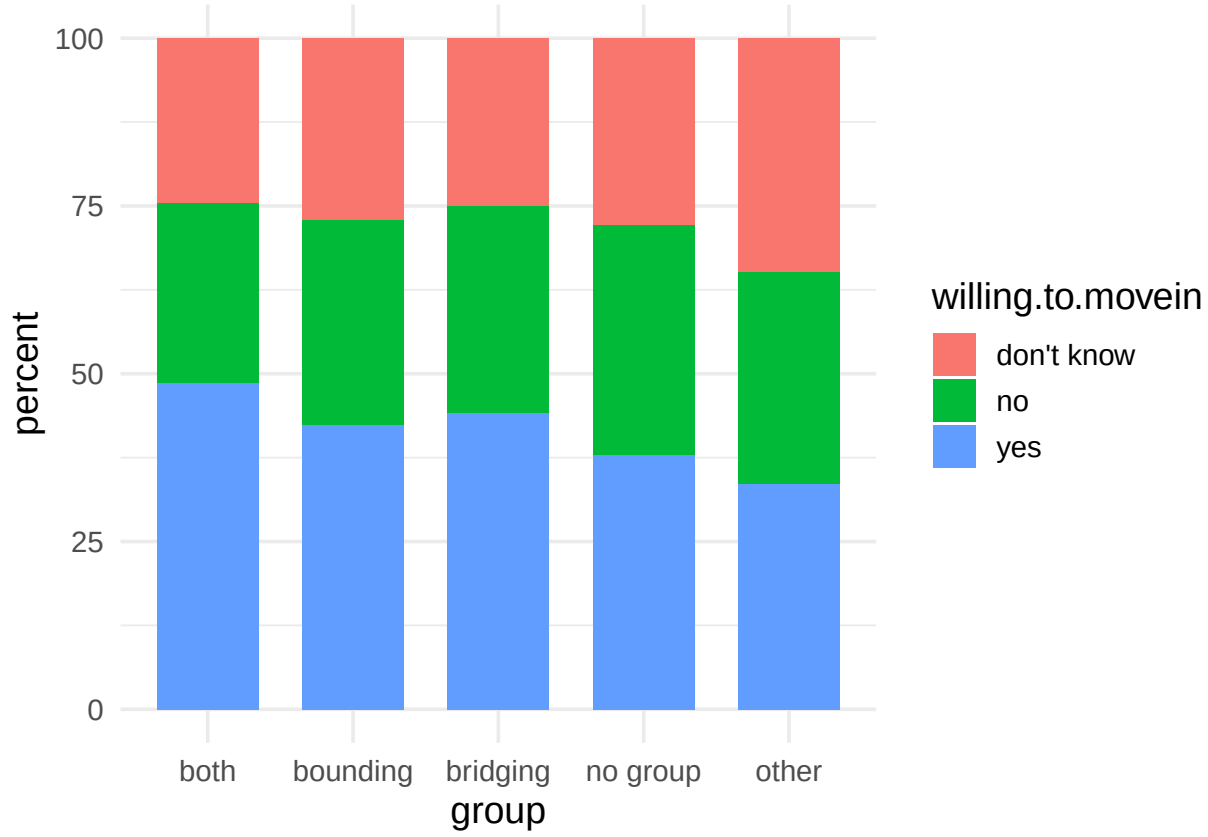
```
library(dplyr)
```

```
cbPalette <- c("#e61212", "#ffb300", "#22ff00", "#0015ff", "#00fbff")
d2 <- dat %>%
  group_by(group, willing.to.movein) %>%
```

```
  summarise(count = n()) %>%
  mutate(perc = count/sum(count))
```

```
## `summarise()` has grouped output by 'group'. You can override using the `.groups` argument.
```

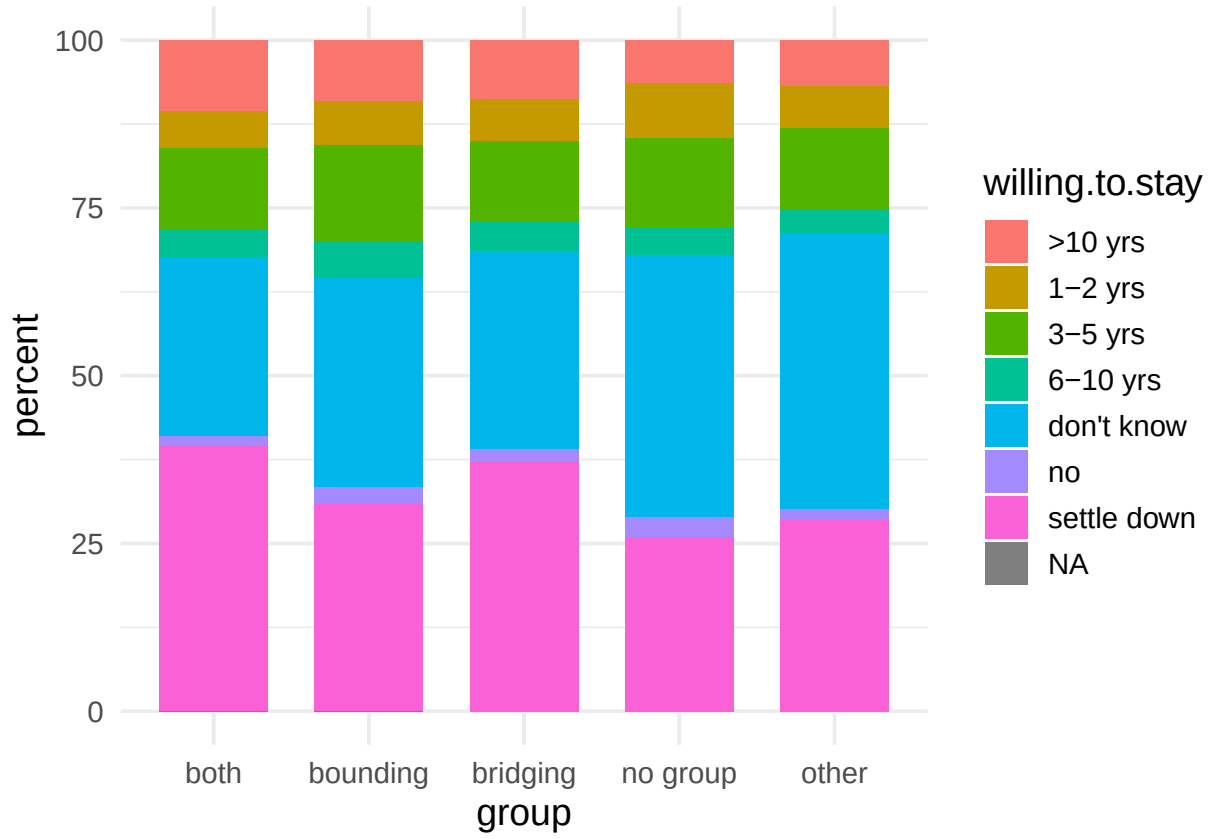
```
ggplot(d2, aes(x = factor(group), y = perc*100, fill = factor(willing.to.movein))) +
  geom_bar(stat="identity", width = 0.7) +
  labs(x = "group", y = "percent", fill = "willing.to.movein") +
  theme_minimal(base_size = 14)
```




```
cbPalette <- c("#e61212", "#ffb300", "#22ff00", "#0015ff", "#00fbff")
d2 <- dat %>%
  group_by(group, willing.to.stay) %>%
  summarise(count = n()) %>%
  mutate(perc = count/sum(count))
```

```
## `summarise()` has grouped output by 'group'. You can override using the `.groups` argument.
```

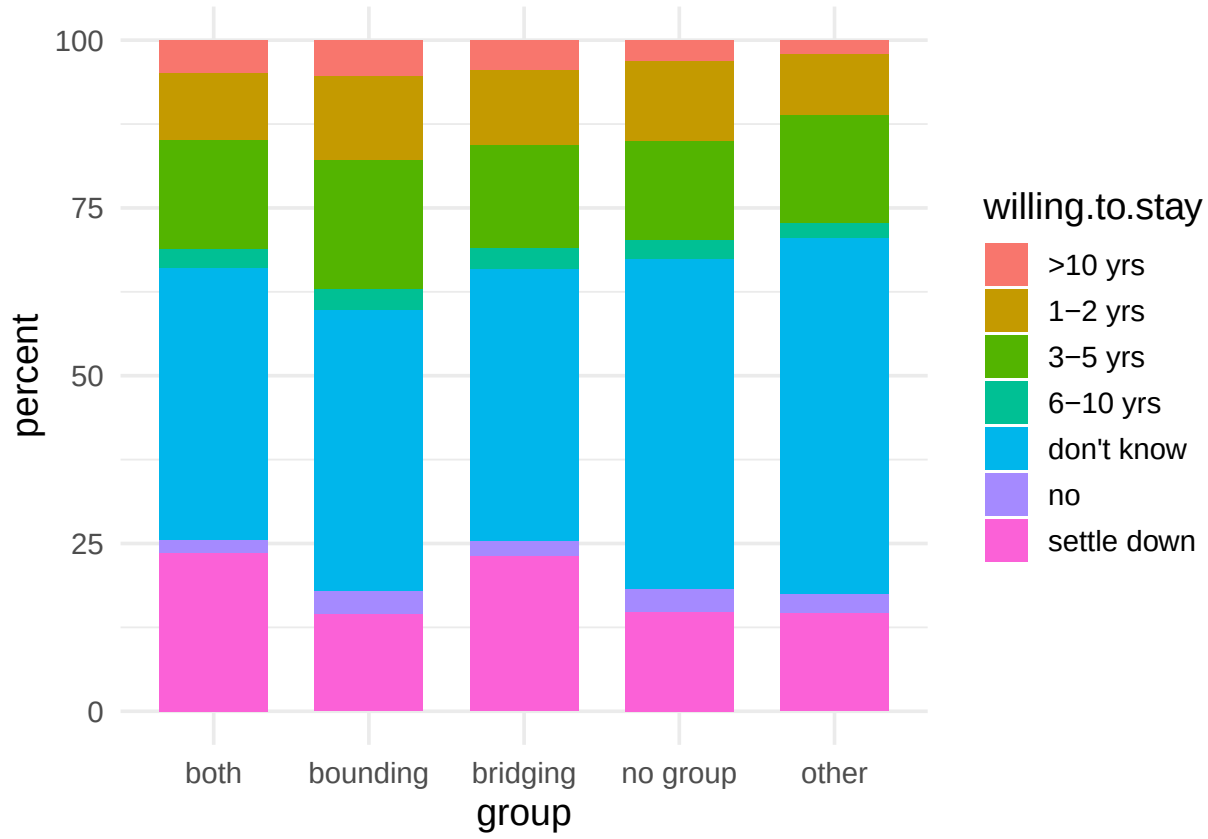
```
ggplot(d2, aes(x = factor(group), y = perc*100, fill = factor(willing.to.stay))) +
  geom_bar(stat="identity", width = 0.7) +
  labs(x = "group", y = "percent", fill = "willing.to.stay") +
  theme_minimal(base_size = 14)
```




```
unmarried <- dat[dat$marriage %in% c("unmarried","divorced","widowed"), ]
d2.unmarried <- unmarried %>%
  group_by(group, willing.to.stay) %>%
  summarise(count = n()) %>%
  mutate(perc = count/sum(count))
```

```
## `summarise()` has grouped output by 'group'. You can override using the `.groups` argument.
```

```
ggplot(d2.unmarried, aes(x = factor(group), y = perc*100, fill = factor(willing.to.stay))) +
  geom_bar(stat="identity", width = 0.7) +
  labs(x = "group", y = "percent", fill = "willing.to.stay") +
  theme_minimal(base_size = 14)
```




```
dat$willing.to.stay <- relevel(dat$willing.to.stay, ref = "no")
dat$group <- relevel(dat$group, ref = "no group")
dat$job <- relevel(dat$job, ref = "unstable job")
dat$education.group <- relevel(as.factor(dat$education.group), ref = "low")
dat$ethnicity <- relevel(as.factor(dat$ethnicity), ref = "1")
```

```
mod1 <- mblogit(formula= willing.to.stay ~ group, data=dat)
```

```
##
## Iteration 1 - deviance = 108455.9 - criterion = 0.3173676
## Iteration 2 - deviance = 107423.2 - criterion = 0.009613114
## Iteration 3 - deviance = 107409.3 - criterion = 0.0001293702
## Iteration 4 - deviance = 107409.3 - criterion = 3.901142e-07
## Iteration 5 - deviance = 107409.3 - criterion = 2.38617e-11
## converged
```

```
summary(mod1)
```

```
##
## Call:
## mblogit(formula = willing.to.stay ~ group, data = dat)
##
## Equation for >10 yrs vs no:
##                Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)    0.76303    0.05285  14.438  < 2e-16 ***
## groupboth      1.14503    0.16050   7.134 9.74e-13 ***
## groupbounding  0.62327    0.18891   3.299 0.000969 ***
## groupbridging  0.75856    0.09508   7.978 1.49e-15 ***
## groupother     0.74843    0.29011   2.580 0.009886 **
##
## Equation for 1-2 yrs vs no:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.03369    0.05082  20.342   <2e-16 ***
## groupboth      0.23043    0.16802   1.371   0.1702
## groupbounding  0.03825    0.19470   0.196   0.8442
## groupbridging  0.16256    0.09623   1.689   0.0912 .
## groupother     0.36913    0.29265   1.261   0.2072
##
## Equation for 3-5 yrs vs no:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.49920    0.04827  31.058  < 2e-16 ***
## groupboth      0.52895    0.15799   3.348 0.000814 ***
## groupbounding  0.34801    0.18109   1.922 0.054641 .
## groupbridging  0.32618    0.09102   3.584 0.000339 ***
## groupother     0.56344    0.27834   2.024 0.042939 *
##
## Equation for 6-10 yrs vs no:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.31446    0.05741   5.478 4.31e-08 ***
## groupboth      0.69349    0.17491   3.965 7.35e-05 ***
## groupbounding  0.54776    0.20180   2.714  0.00664 **
## groupbridging  0.52400    0.10316   5.080 3.78e-07 ***
## groupother     0.56101    0.31263   1.794  0.07274 .
##
## Equation for don't know vs no:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.57762    0.04527  56.938  < 2e-16 ***
## groupboth      0.24851    0.15243   1.630  0.10302
## groupbounding  0.04820    0.17398   0.277  0.78177
## groupbridging  0.15007    0.08668   1.731  0.08339 .
## groupother     0.71327    0.26683   2.673  0.00751 **
##
## Equation for settle down vs no:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.17020    0.04607  47.109  < 2e-16 ***
## groupboth      1.05187    0.15139   6.948  3.7e-12 ***
## groupbounding  0.44605    0.17424   2.560  0.01047 *
## groupbridging  0.79217    0.08669   9.138  < 2e-16 ***
## groupother     0.76011    0.26898   2.826  0.00471 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Null Deviance:     131100
## Residual Deviance: 107400
## Number of Fisher Scoring iterations:  5
## Number of observations:  33695
##   (2 observations deleted due to missingness)
```

```
mod2 <- mblogit(formula= willing.to.stay ~ group + city_inclusion + money.left + education.group, data=
```

```
##
## Iteration 1 - deviance = 104539.1 - criterion = 0.291938
## Iteration 2 - deviance = 102867.3 - criterion = 0.016252
## Iteration 3 - deviance = 102813.7 - criterion = 0.0005212565
## Iteration 4 - deviance = 102813.2 - criterion = 4.125733e-06
## Iteration 5 - deviance = 102813.2 - criterion = 1.063697e-09
## converged
```

```
summary(mod2)
```

```
##
## Call:
## mblogit(formula = willing.to.stay ~ group + city_inclusion +
##     money.left + education.group, data = dat)
##
## Equation for >10 yrs vs no:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -0.4054     0.1134  -3.575 0.000351 ***
## groupboth               0.8896     0.1628   5.463 4.67e-08 ***
## groupbounding           0.5752     0.1928   2.984 0.002845 **
## groupbridging           0.5735     0.0968   5.925 3.13e-09 ***
## groupother              0.6899     0.2914   2.368 0.017907 *
## city_inclusion          0.2595     0.0138  18.804  < 2e-16 ***
## money.left              0.1952     0.0492   3.967 7.29e-05 ***
## education.grouphigh     0.4184     0.1872   2.235 0.025398 *
## education.groupmiddle   0.3973     0.1154   3.444 0.000574 ***
##
## Equation for 1-2 yrs vs no:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.55851    0.10505   5.317 1.06e-07 ***
## groupboth              0.12409    0.16965   0.731   0.4645
## groupbounding          0.03024    0.19745   0.153   0.8783
## groupbridging          0.09321    0.09744   0.957   0.3388
## groupother             0.33260    0.29301   1.135   0.2563
## city_inclusion         0.10363    0.01345   7.705 1.31e-14 ***
## money.left             0.05960    0.05162   1.155   0.2482
## education.grouphigh    0.14494    0.18913   0.766   0.4435
## education.groupmiddle  0.26476    0.11011   2.405   0.0162 *
##
## Equation for 3-5 yrs vs no:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.56866    0.10214   5.567 2.59e-08 ***
## groupboth              0.34588    0.15978   2.165  0.03041 *
## groupbounding          0.30674    0.18435   1.664  0.09613 .
## groupbridging          0.20150    0.09237   2.181  0.02915 *
## groupother             0.50524    0.27905   1.811  0.07021 .
## city_inclusion         0.16054    0.01279  12.553  < 2e-16 ***
## money.left             0.07771    0.04897   1.587  0.11256
## education.grouphigh    0.50615    0.17924   2.824  0.00475 **
## education.groupmiddle  0.57527    0.10652   5.401 6.63e-08 ***
```

```
##
## Equation for 6-10 yrs vs no:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.67468    0.12694  -5.315 1.07e-07 ***
## groupboth              0.48181    0.17701   2.722 0.006491 **
## groupbounding          0.50492    0.20508   2.462 0.013814 *
## groupbridging          0.37422    0.10478   3.572 0.000355 ***
## groupother             0.50338    0.31348   1.606 0.108324
## city_inclusion         0.20078    0.01506  13.337  < 2e-16 ***
## money.left             0.13907    0.05315   2.616 0.008884 **
## education.grouphigh    0.44041    0.20568   2.141 0.032253 *
## education.groupmiddle  0.47184    0.12893   3.660 0.000253 ***
##
## Equation for don't know vs no:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            1.82192    0.09207  19.789  < 2e-16 ***
## groupboth              0.11394    0.15408   0.739   0.4596
## groupbounding          0.03584    0.17718   0.202   0.8397
## groupbridging          0.05212    0.08791   0.593   0.5533
## groupother             0.65933    0.26743   2.465   0.0137 *
## city_inclusion         0.14071    0.01200  11.730  < 2e-16 ***
## money.left            -0.07491    0.04730  -1.584   0.1133
## education.grouphigh    0.39562    0.16880   2.344   0.0191 *
## education.groupmiddle  0.42916    0.09726   4.412 1.02e-05 ***
##
## Equation for settle down vs no:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.12778    0.10120  -1.263   0.2067
## groupboth              0.60411    0.15453   3.909 9.26e-05 ***
## groupbounding          0.34388    0.17967   1.914   0.0556 .
## groupbridging          0.43085    0.08907   4.837 1.32e-06 ***
## groupother             0.67939    0.27181   2.499   0.0124 *
## city_inclusion         0.39801    0.01254  31.742  < 2e-16 ***
## money.left             0.20256    0.04708   4.302 1.69e-05 ***
## education.grouphigh    1.54870    0.17155   9.028  < 2e-16 ***
## education.groupmiddle  0.84222    0.10397   8.101 5.47e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Null Deviance:     131100
## Residual Deviance: 102800
## Number of Fisher Scoring iterations:  5
## Number of observations:  33688
##   (9 observations deleted due to missingness)
```

```
write_csv(dat,"processed_data.csv")
```