Survey_data_analyses

Christoph Völtzke

2023-01-26

```
library(readr)
library(survey)
library(tidyverse)
library(covidregionaldata)
library(socialmixr)
library(ggplot2)
```

```
complete.data <- read_csv("Data/Data_processed/full_data.csv")
load("Data/uk_pop.Rdata")
```

Poststratify on age, gender, and household type

These are the only variables in common that are eligilbe for the use for post stratification purposes.

```
# switching some var names for plots
complete.data <-complete.data %>%
  mutate(part_gender = case_when(part_gender == "M" ~ "Male",
                                 part_gender == "F" ~ "Female"),
         hh_type = case_when(hh_type == 1 ~ "Not alone",
                             hh_type == 0 ~ "Alone"),
         senior = case_when(senior == 1 ~ "Senior",
                             senior == 0 ~ "No senior")
  )
```

```
# define survey design without weights
svy.unweighted <- svydesign(ids=~part_id, data=complete.data)

# get the postratification weights by age, gender, and household type
uk_pop <- uk_pop %>%
  mutate(part_gender = case_when(part_gender == "M" ~ "Male",
                                 part_gender == "F" ~ "Female"),
         hh_type = case_when(hh_type == 1 ~ "Not alone",
                             hh_type == 0 ~ "Alone")) %>%
  group_by(part_gender, part_age_group, hh_type) %>%
  slice(rep(1:n(), each = pop_estimate))
# create contingency table
pop_data.dist <- xtabs(~part_gender+part_age_group+hh_type, data=uk_pop)
# define post-stratification survey design
svy.poststrat <- postStratify(design=svy.unweighted,
                              strata=~part_gender+part_age_group+hh_type,
                              population=pop_data.dist)
```

Running analyses with the svy weights

```
glm.fit <- svyglm(contacts ~ part_age_group + part_gender + factor(hh_type) + date, design=svy.poststra
glm.gov <- svyglm(contacts ~ stringency_index + cases_new + date, design=svy.poststrat)
glm.age <- svyglm(contacts ~ part_age_group + date, design=svy.poststrat)
glm.gender <- svyglm(contacts ~ part_gender + date, design=svy.poststrat)
glm.hhtype <- svyglm(contacts ~ factor(hh_type) + date, design=svy.poststrat)
glm.location <- svyglm(contacts ~ factor(location) + date, design=svy.poststrat)
glm.periods <- svyglm(contacts ~ factor(periods) + date, design=svy.poststrat)
glm.senior.gen <- svyglm(contacts ~ factor(senior)*part_gender + date, design=svy.poststrat)
glm.senior.hh <- svyglm(contacts ~ factor(senior)*factor(hh_type) + date, design=svy.poststrat)
```

```
summary(glm.fit)
```

```
##
## Call:
## svyglm(formula = contacts ~ part_age_group + part_gender + factor(hh_type) +
##     date, design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               18.3644349  6.7594017   2.717 0.006595 **
## part_age_group25-44       -0.0506658  0.0642934  -0.788 0.430681
## part_age_group45-64       -0.0391849  0.0634126  -0.618 0.536624
## part_age_group65 and over -0.6782012  0.0642294 -10.559  < 2e-16 ***
## part_genderMale           -0.1046087  0.0272857  -3.834 0.000127 ***
## factor(hh_type)Not alone   0.2093200  0.0500848   4.179 2.93e-05 ***
## date                      -0.0008540  0.0003669  -2.328 0.019930 *
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for gaussian family taken to be 3.651694)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.gov)
```

```
##
## Call:
## svyglm(formula = contacts ~ stringency_index + cases_new + date,
##     design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.236e+01  1.449e+01   4.995 5.94e-07 ***
## stringency_index  1.569e-03  5.061e-03   0.310    0.757
## cases_new        -9.533e-05  1.593e-05  -5.985 2.20e-09 ***
## date             -3.785e-03  7.738e-04  -4.892 1.00e-06 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for gaussian family taken to be 3.733398)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.age)
```

```
##
## Call:
## svyglm(formula = contacts ~ part_age_group + date, design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               18.6484616  6.7638951   2.757  0.00584 **
## part_age_group25-44       -0.0589219  0.0644717  -0.914  0.36077
## part_age_group45-64       -0.0624903  0.0637708  -0.980  0.32714
## part_age_group65 and over -0.7276061  0.0644612 -11.288  < 2e-16 ***
## date                      -0.0008614  0.0003674  -2.345  0.01906 *
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for gaussian family taken to be 3.659552)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.gender)
```

```
##
## Call:
## svyglm(formula = contacts ~ part_gender + date, design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     22.9129373  6.8698989   3.335 0.000854 ***
## part_genderMale -0.0825463  0.0273643  -3.017 0.002559 **
## date            -0.0011024  0.0003731  -2.954 0.003138 **
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
##
## (Dispersion parameter for gaussian family taken to be 3.740464)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.hhtype)
```

```
##
## Call:
## svyglm(formula = contacts ~ factor(hh_type) + date, design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              20.730645   6.873921   3.016  0.00257 **
## factor(hh_type)Not alone  0.382238   0.048714   7.847 4.46e-15 ***
## date                     -0.001004   0.000373  -2.690  0.00714 **
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for gaussian family taken to be 3.723615)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.location)
```

```
##
## Call:
## svyglm(formula = contacts ~ factor(location) + date, design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                20.8653744  6.7407852   3.095  0.00197 **
## factor(location)Leisure     1.0374593  0.5020713   2.066  0.03881 *
## factor(location)Otherplace -0.3696404  0.0441937  -8.364  < 2e-16 ***
## factor(location)School     -0.0489686  0.2992237  -0.164  0.87001
## factor(location)Transport   0.4495475  0.2640685   1.702  0.08870 .
## factor(location)Work        1.4319049  0.1738255   8.238  < 2e-16 ***
## date                       -0.0009938  0.0003661  -2.714  0.00664 **
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for gaussian family taken to be 3.660887)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.senior.gen)
```

```
##
## Call:
## svyglm(formula = contacts ~ factor(senior) * part_gender + date,
##     design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          19.7250320  6.7813411   2.909  0.00363 **
## factor(senior)Senior                 -0.6688409  0.0436210 -15.333  < 2e-16 ***
## part_genderMale                      -0.0973806  0.0325295  -2.994  0.00276 **
## date                                 -0.0009201  0.0003683  -2.498  0.01250 *
## factor(senior)Senior:part_genderMale -0.0227071  0.0576189  -0.394  0.69352
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for gaussian family taken to be 3.657231)
##
## Number of Fisher Scoring iterations: 2
```

```
summary(glm.senior.hh)
```

```
##
## Call:
## svyglm(formula = contacts ~ factor(senior) * factor(hh_type) +
##     date, design = svy.poststrat)
##
## Survey design:
## postStratify(design = svy.unweighted, strata = ~part_gender +
##     part_age_group + hh_type, population = pop_data.dist)
##
## Coefficients:
##                                                 Estimate Std. Error t value
## (Intercept)                                   18.4555567  6.8037751   2.713
## factor(senior)Senior                          -0.2968453  0.0920502  -3.225
## factor(hh_type)Not alone                       0.3861692  0.0721309   5.354
## date                                          -0.0008725  0.0003691  -2.364
## factor(senior)Senior:factor(hh_type)Not alone -0.4322277  0.0965002  -4.479
##                                               Pr(>|t|)
## (Intercept)                                    0.00668 **
## factor(senior)Senior                           0.00126 **
## factor(hh_type)Not alone                      8.70e-08 ***
## date                                           0.01809 *
## factor(senior)Senior:factor(hh_type)Not alone 7.53e-06 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
```

```
## (Dispersion parameter for gaussian family taken to be 3.64918)
##
## Number of Fisher Scoring iterations: 2
```

Datasets needed for plotting with/wo Survey design

These seperated data sets are all needed to obtain the plots!

```r
# contacts (with and w/o poststratification)
data.wgt <- svyby(~contacts, ~weeks, design=svy.poststrat, svymean)
data.wgt <- data.wgt %>%
  mutate(wgt = 0)

data.non.wgt <- svyby(~contacts, ~weeks, design=svy.unweighted, svymean)
data.non.wgt <- data.non.wgt %>%
  mutate(wgt = 1)

data.weeks <- rbind(data.wgt, data.non.wgt)
```

```r
# contacts by age group (with and w/o poststratification)
data.age.wgt <- svyby(~contacts, ~part_age_group+weeks, design=svy.poststrat, svymean)
data.age.wgt <- data.age.wgt %>%
  mutate(wgt = 0)

data.age.non.wgt <- svyby(~contacts, ~part_age_group+weeks, design=svy.unweighted, svymean)
data.age.non.wgt <- data.age.non.wgt %>%
  mutate(wgt = 1)

data.age.weeks <- rbind(data.age.wgt, data.age.non.wgt)
```

```r
# contacts by gender (with and w/o poststratification)
data.gender.wgt <- svyby(~contacts, ~part_gender+weeks,design=svy.poststrat, svymean)
data.gender.wgt <- data.gender.wgt %>%
  mutate(wgt = 0)

data.gender.non.wgt <- svyby(~contacts, ~part_gender+weeks, design=svy.unweighted, svymean)
data.gender.non.wgt <- data.gender.non.wgt %>%
  mutate(wgt = 1)

data.gender.weeks <- rbind(data.gender.wgt, data.gender.non.wgt)
```

```r
# contacts by household type (with and w/o poststratification)
data.hhtype.wgt <- svyby(~contacts, ~hh_type+weeks, design=svy.poststrat, svymean)
data.hhtype.wgt <- data.hhtype.wgt %>%
  mutate(wgt = 0)

data.hhtype.non.wgt <- svyby(~contacts, ~hh_type+weeks, design=svy.unweighted, svymean)
data.hhtype.non.wgt <- data.hhtype.non.wgt %>%
  mutate(wgt = 1)

data.hhtype.weeks <- rbind(data.hhtype.wgt, data.hhtype.non.wgt)
```

```r
# contacts of 65 year olds and over by gender (with and w/o poststratification)
data.seniors.wgt <- svyby(~contacts, ~senior+part_gender+weeks, design=svy.poststrat, svymean)
data.seniors.wgt <- data.seniors.wgt %>%
  mutate(wgt = 0)

data.seniors.non.wgt <- svyby(~contacts, ~senior+part_gender+weeks, design=svy.unweighted, svymean)
data.seniors.non.wgt <- data.seniors.non.wgt %>%
  mutate(wgt = 1)

data.seniors.weeks <- rbind(data.seniors.wgt, data.seniors.non.wgt)
```

```r
# contacts of 65 year olds and over by household type (with and w/o
# poststratification)
data.seniors.hh.wgt <- svyby(~contacts, ~senior+hh_type+weeks, design=svy.poststrat, svymean)
data.seniors.hh.wgt <- data.seniors.hh.wgt %>%
  mutate(wgt = 0)

data.seniors.hh.non.wgt <- svyby(~contacts, ~senior+hh_type+weeks, design=svy.unweighted, svymean)
data.seniors.hh.non.wgt <- data.seniors.hh.non.wgt %>%
  mutate(wgt = 1)

data.seniors.hh <- rbind(data.seniors.hh.wgt, data.seniors.hh.non.wgt)
```

```r
# cases per week
cases.weeks <- subset(complete.data, select=c(weeks, cases_new))
cases.weeks <- cases.weeks %>%
  group_by(weeks) %>%
  mutate(cases = mean(cases_new)) %>%
  distinct(weeks, .keep_all=TRUE)
```

```r
source("Functions/plotting_post_strat.R")
source("Functions/plotting_with_Covid.R")
```

```r
plotting_post(data.weeks,data.weeks$wgt,"Figure 1: Contacts over Time")
```

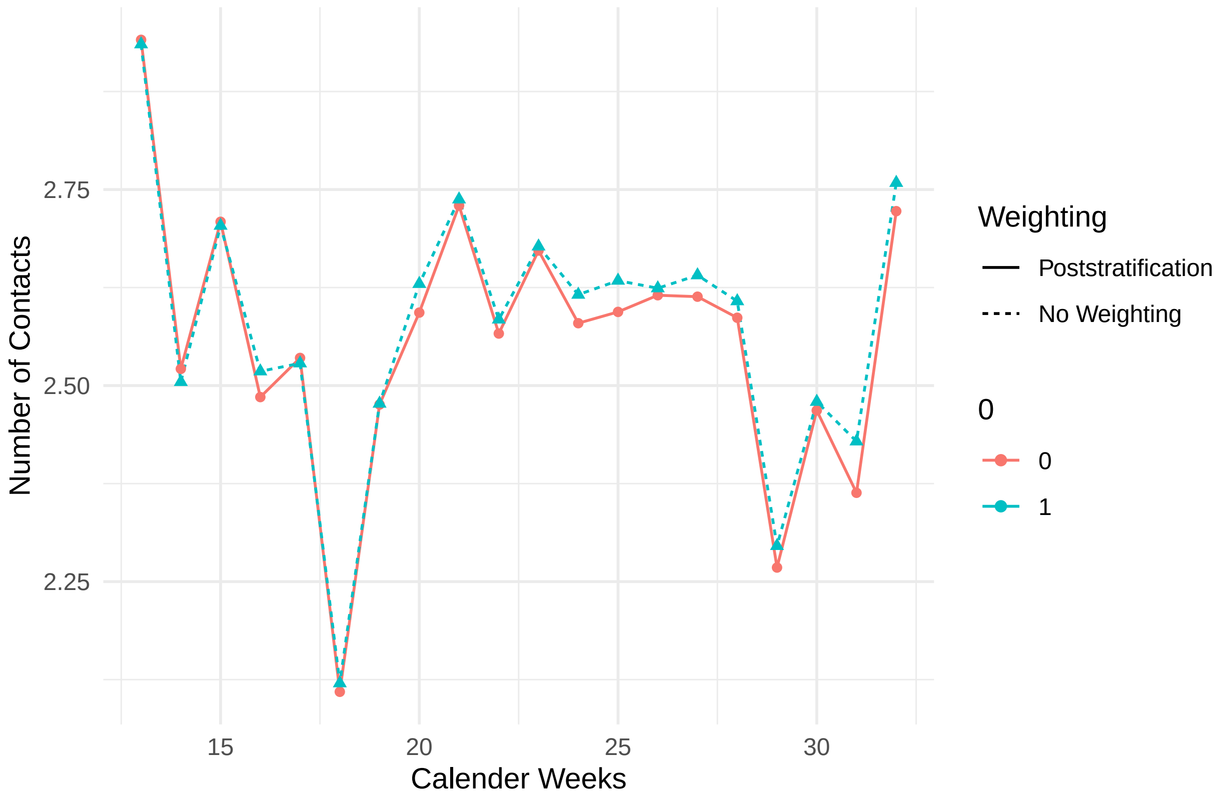
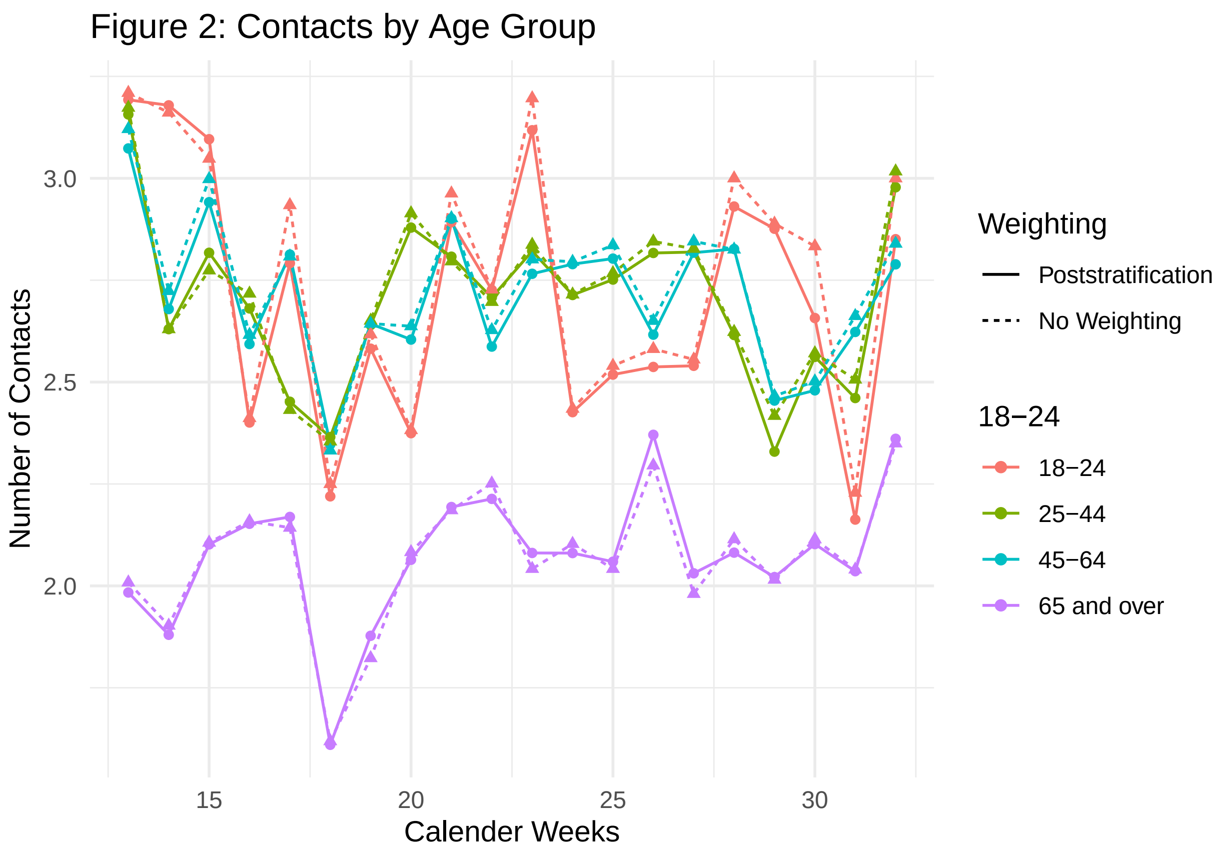


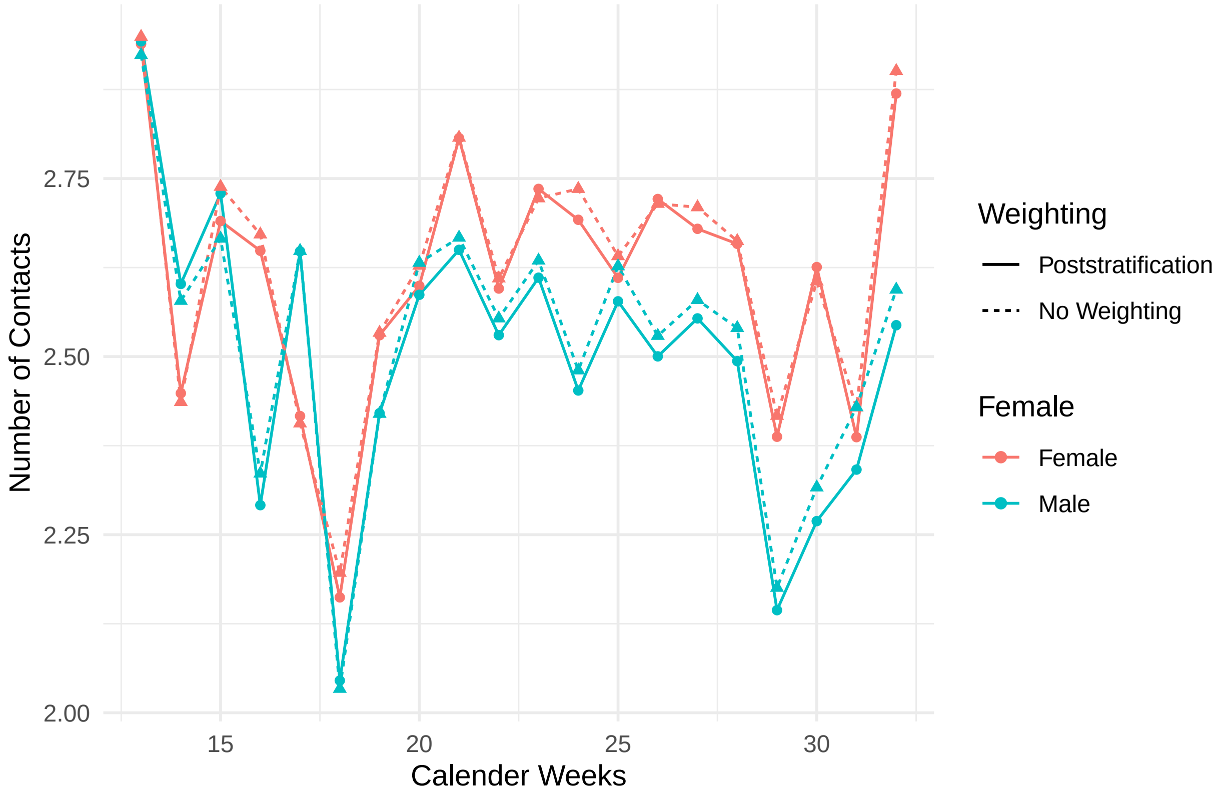
Figure 1: Contacts over Time

```
plotting_post(data.age.weeks,data.age.weeks$part_age_group,"Figure 2: Contacts by Age Group")
```



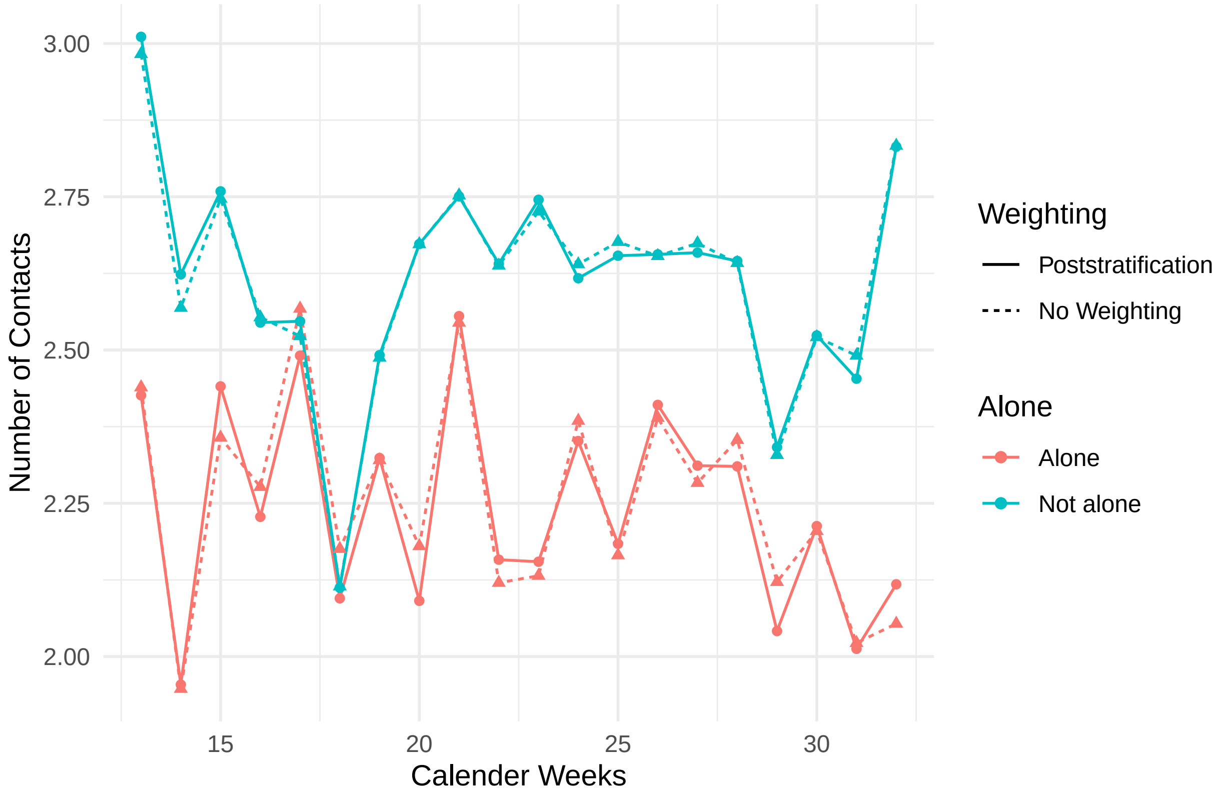
```
plotting_post(data.gender.weeks,data.gender.weeks$part_gender,"Figure 3: Contacts by Gender")
```

Figure 3: Contacts by Gender



```
plotting_post(data.hhtype.weeks,data.hhtype.weeks$hh_type,"Figure 4: Contacts by Household Typ")
```

Figure 4: Contacts by Household Typ



```
plotting_post(data.seniors.weeks,interaction(factor(data.seniors.weeks$senior), data.seniors.weeks$part_
```

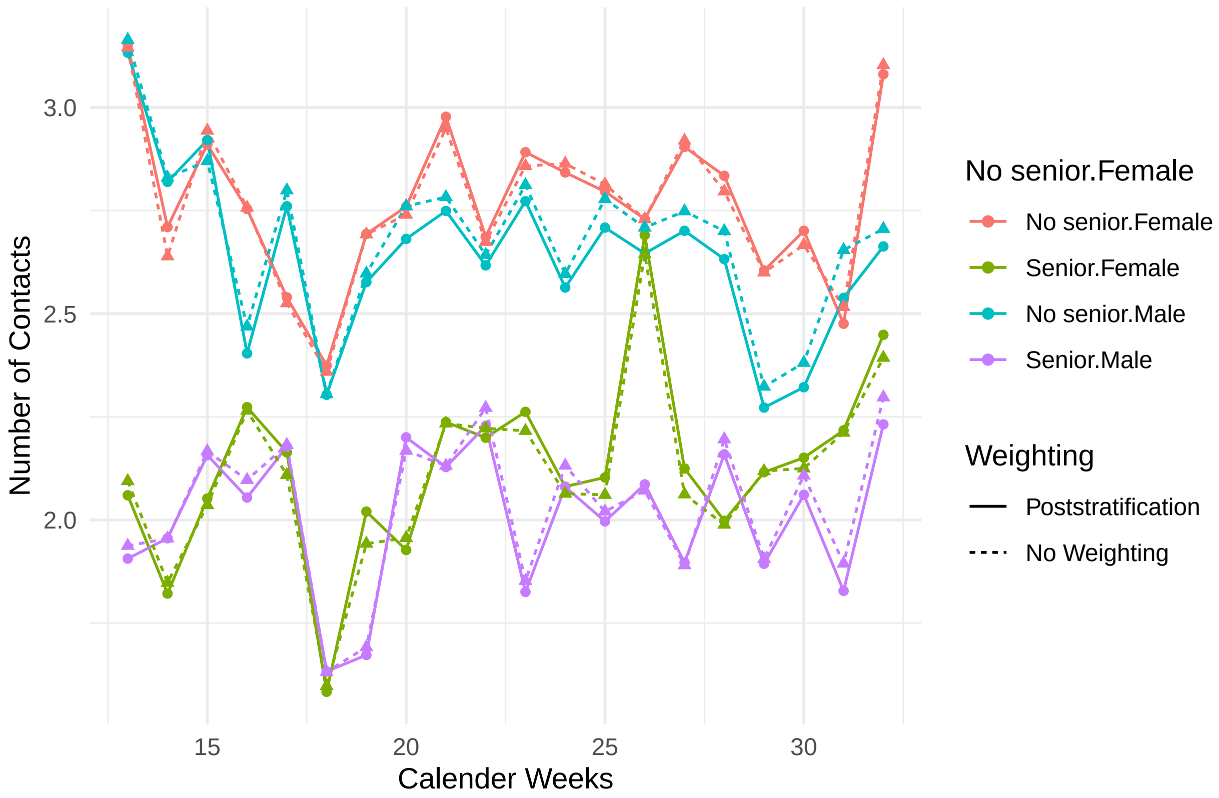
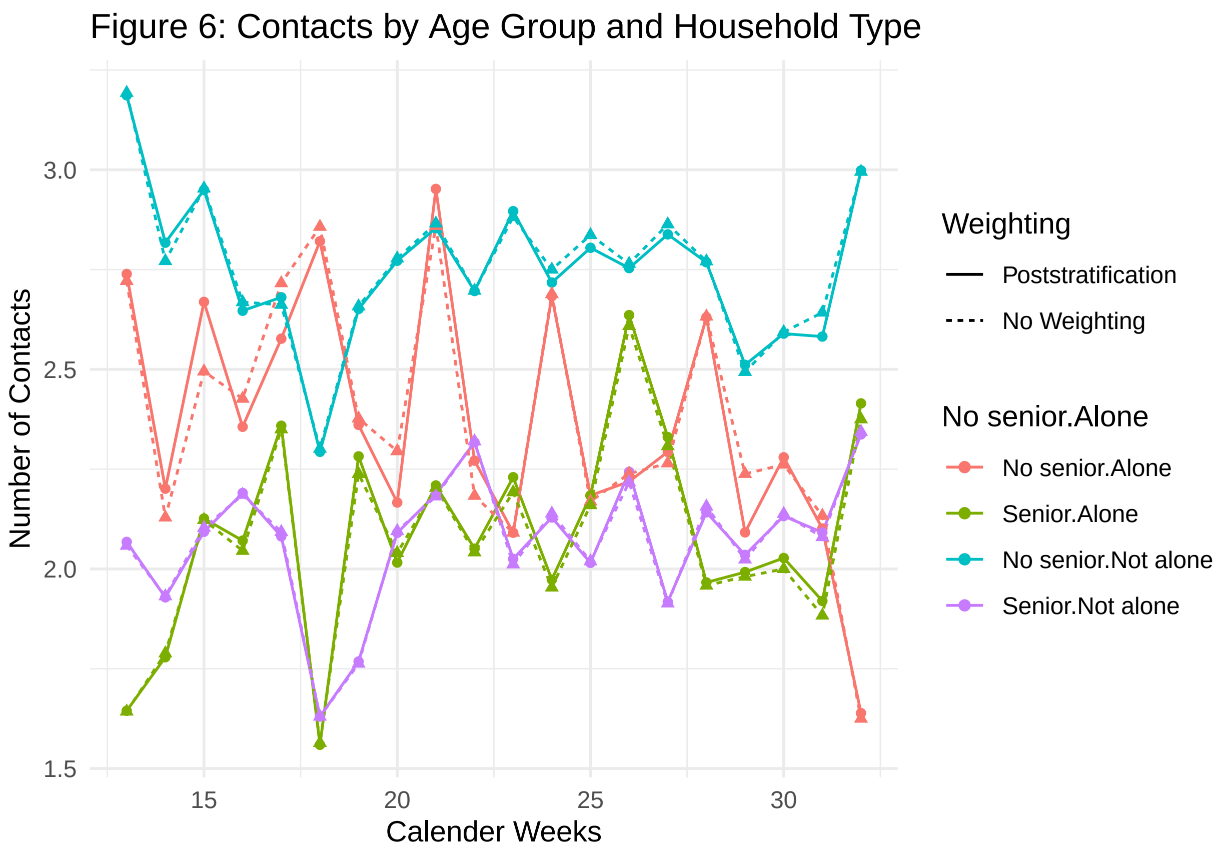


Figure 5: Contacts by Age Group and Gender

```
plotting_post(data.seniors.hh,interaction(factor(data.seniors.hh$senior), data.seniors.hh$hh_type),"Fig
```



```r
plotting_covid(data.weeks,data.weeks$wgt,"Figure 1: Contacts over Time")
```

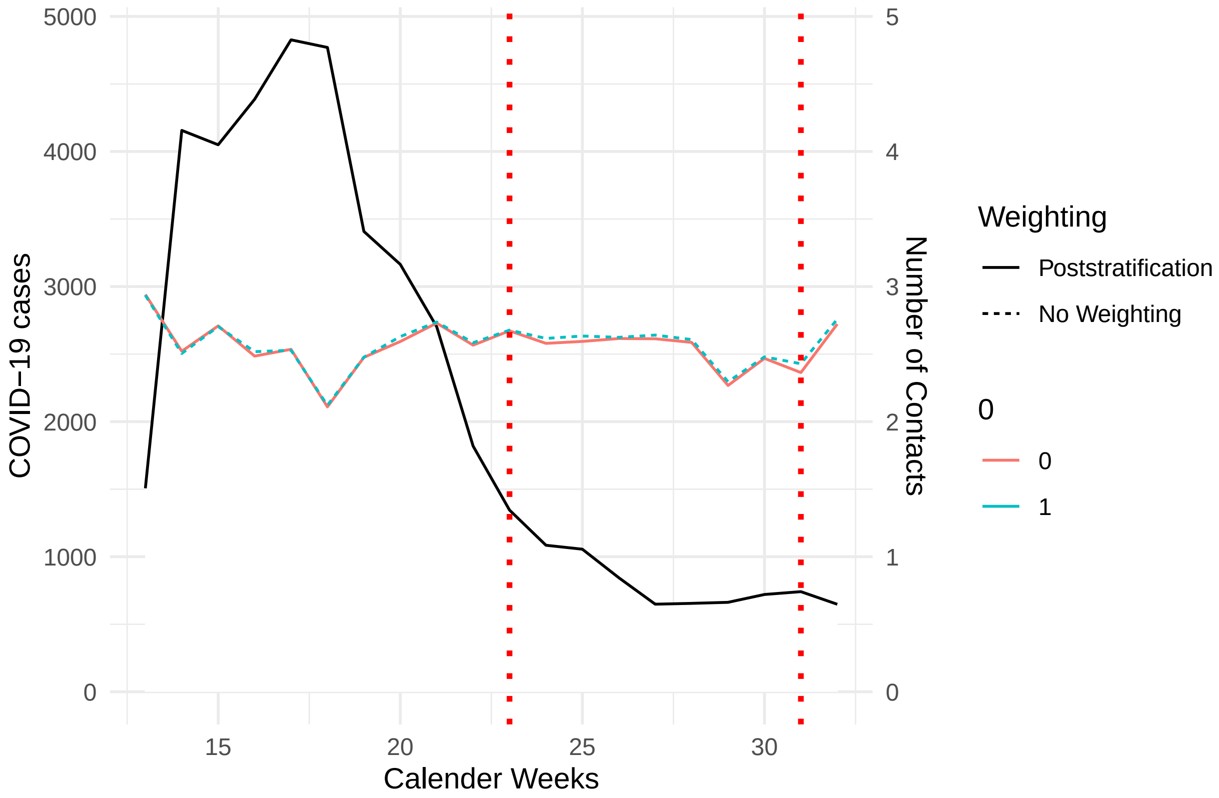
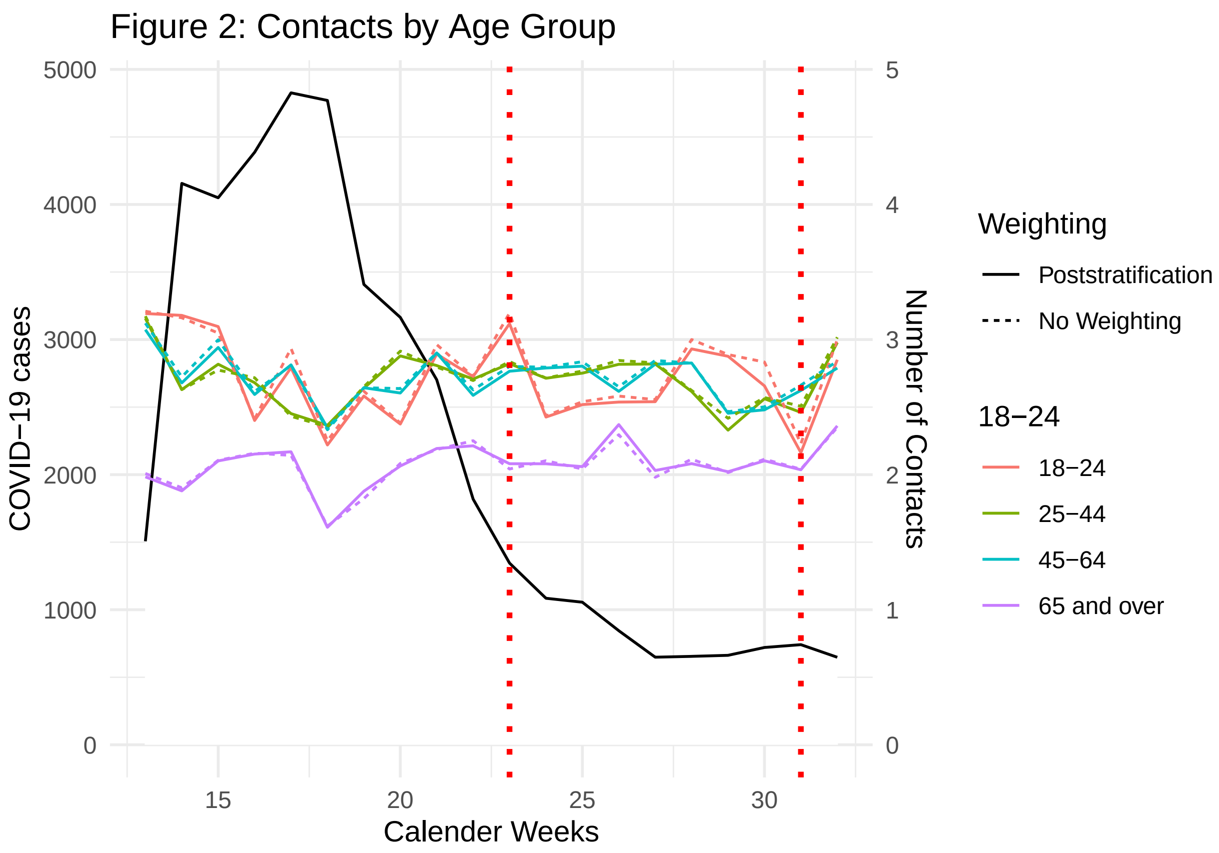
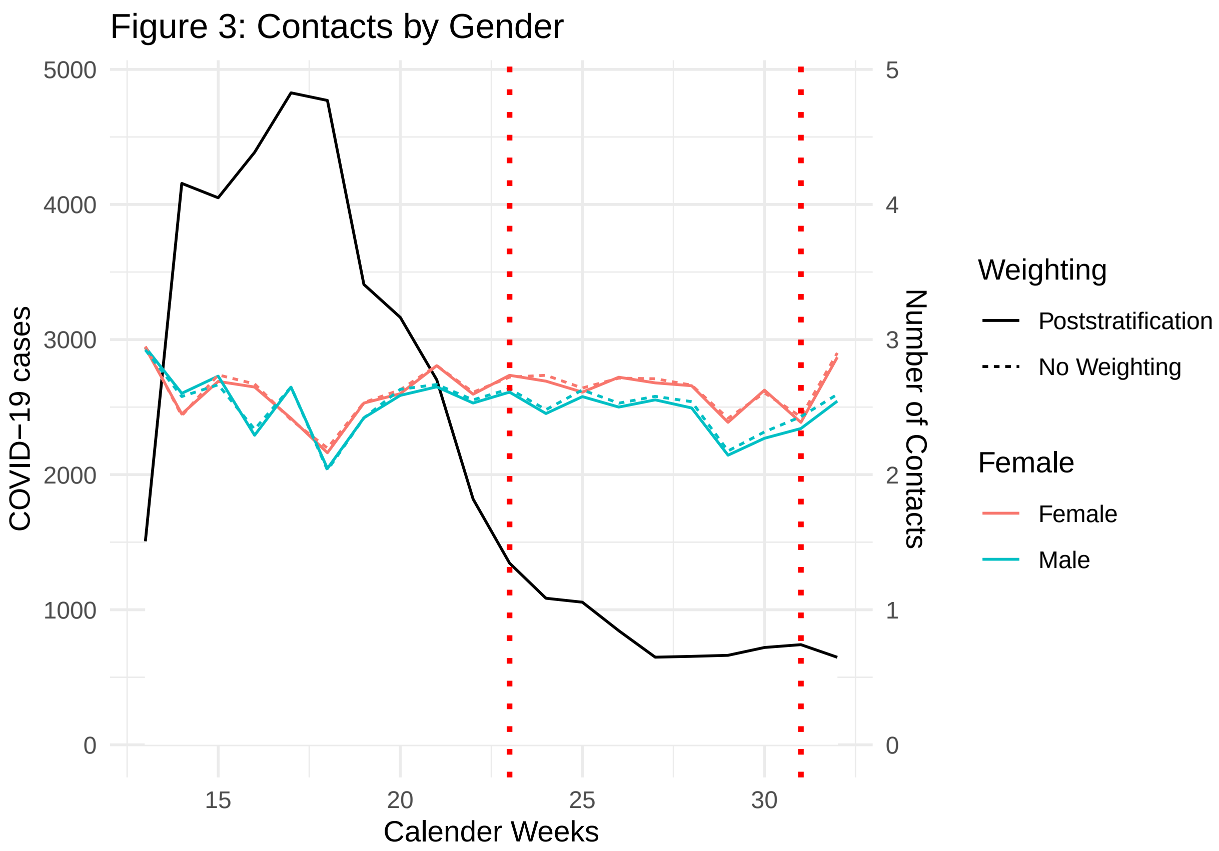


Figure 1: Contacts over Time

```
plotting_covid(data.age.weeks,data.age.weeks$part_age_group,"Figure 2: Contacts by Age Group")
```

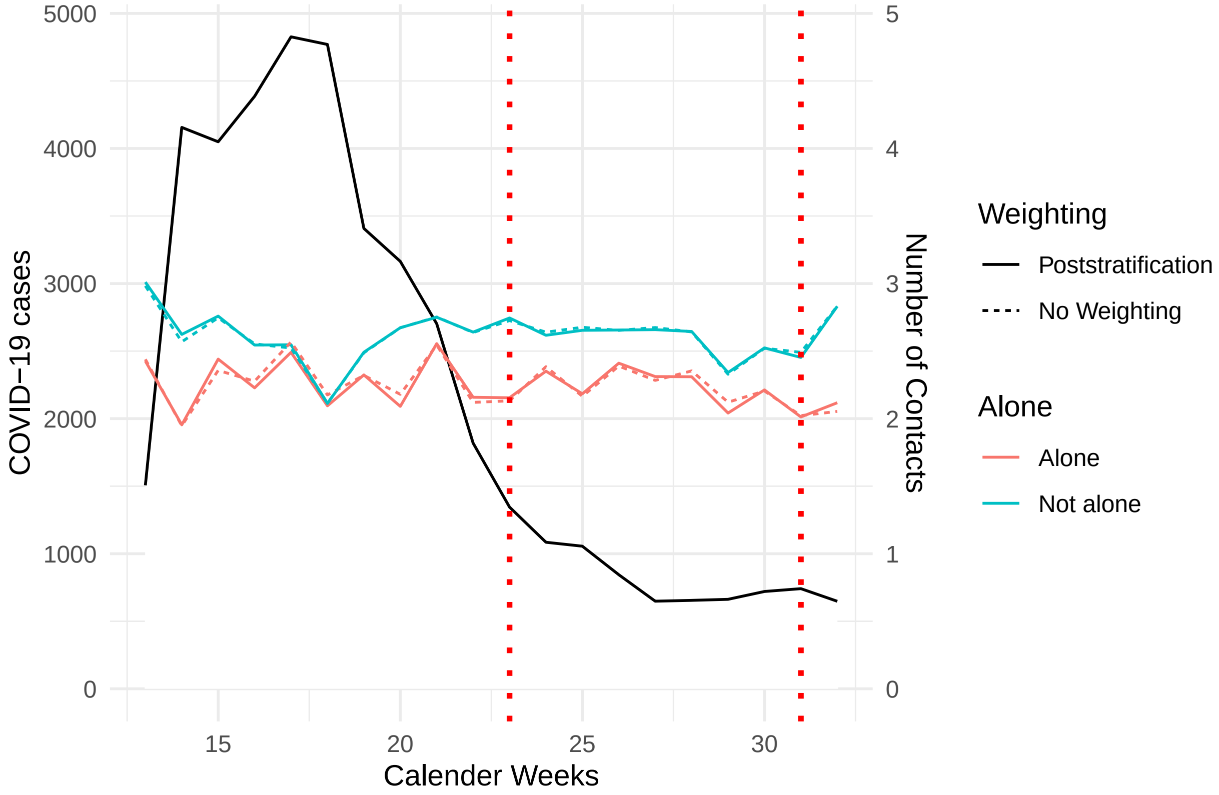


```
plotting_covid(data.gender.weeks,data.gender.weeks$part_gender,"Figure 3: Contacts by Gender")
```



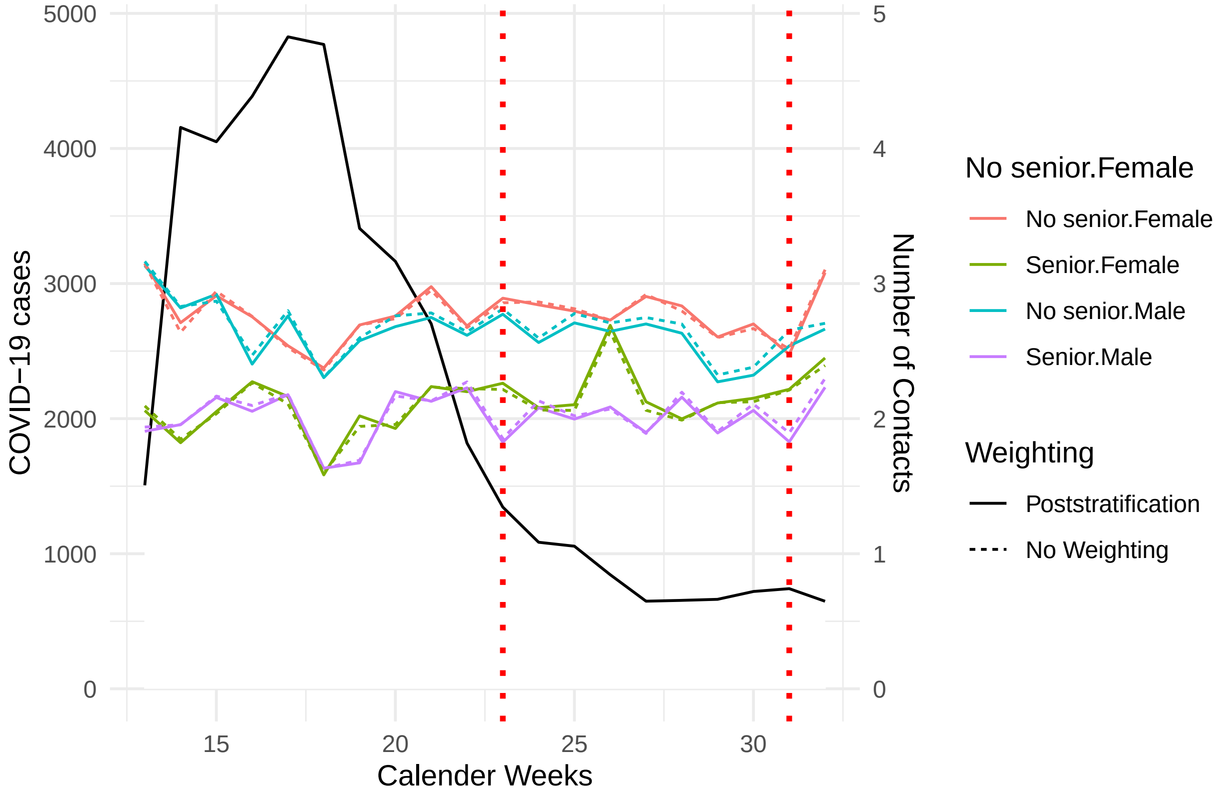
```
plotting_covid(data.hhtype.weeks,data.hhtype.weeks$hh_type,"Figure 4: Contacts by Household Typ")
```

Figure 4: Contacts by Household Typ



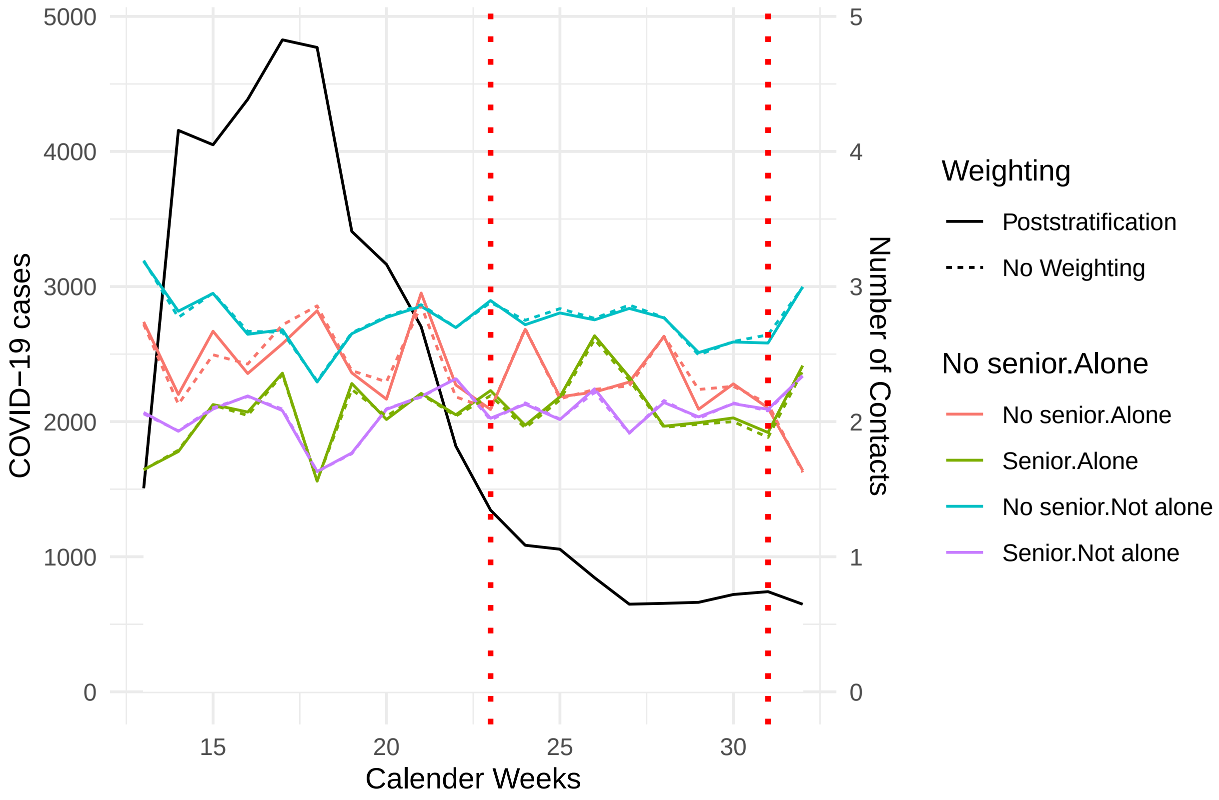
```
plotting_covid(data.seniors.weeks,interaction(factor(data.seniors.weeks$senior), data.seniors.weeks$par
```

Figure 5: Contacts by Age Group and Gender



```
plotting_covid(data.seniors.hh,interaction(factor(data.seniors.hh$senior), data.seniors.hh$hh_type),"Fi
```

Figure 6: Contacts by Age Group and Household Type


COVID−19 cases
Number of Contacts
Calender Weeks
Weighting
Poststratification
No Weighting
No senior.Alone
No senior.Alone
Senior.Alone
No senior.Not alone
Senior.Not alone