830_group_visualCues

Library imports

```
library(data.table)
library(fixest)
library(broom)
library(lfe)
```

```
## Loading required package: Matrix
```

```
library(data.table)
library(ggplot2)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ tibble  3.1.6     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.4     ✓ stringr 1.4.0
## ✓ readr   2.1.1     ✓ forcats 0.5.1
## ✓ purrr   0.3.4
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::between()   masks data.table::between()
## x tidyr::expand()    masks Matrix::expand()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x tidyr::pack()      masks Matrix::pack()
## x purrr::transpose() masks data.table::transpose()
## x tidyr::unpack()    masks Matrix::unpack()
```

```
library(pwr)
library(lfe)
library(modelsummary)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Ta
bles.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
library(dplyr)
library(knitr)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

Abstract: Research question

Today, Bio medical researchers are interested in identifying sophistication levels of people and merging them with safe lifestyles after the outbreak of the novel Corona virus. To interrupt this potential transmission route and to avoid critical commodities, the Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO)(11), and other public health organizations recommend the wearing of face masks or other face coverings by the general public during the ongoing COVID-19 pandemic.

Information on public preferences on masking is generally messy, disparate, and lacks structure. The outcomes of this experiment would aid in understanding the interests and comfort level of people and bring in a more hassle free and undisturbed routine. This randomized controlled trial investigates whether an N95 mask is more preferred to any other mask. This experiment answers the question “What are the significant effects of visual cues in mask preferences when awareness already exists for Covid-19?”

Hypothesis

Below are the null and alternative hypothesis for this experiment.

H0: Visual cues do not affect an individual’s preference of N95 masks over other types of masks.

HA: Visual cues influence an individual’s preference of N95 masks over other types of masks.

Literature review

Visual cues have played a role in solidifying consumer perception in industries such as food and retail. The “first taste is almost always with the eye”. This is especially the case where a food product is sold through its appearance, rather than through its packaging - often the case in sports commercials. Consumers view food commercials and are motivated to place orders. In addition, appearance can have a halo effect which modifies subsequent flavor perception and food acceptability - seeing an ad for a food brand often may push the consumer to believe in brand repute. (1)

Taking this idea, a step forward, visual narratives are promising tools for science and health communication, especially for broad audiences in times of public health crisis, such as during the COVID-19 pandemic. In a day and age, where most consumers are accessing their content visually through social media, it remains an interesting experiment to understand if this behavior could be used for the advantage of the broader public, through visual cues in the form of ads/ informational videos to improve COVID-19 behavior and guideline adherence. COVID-19 focused visual narratives in the form of flashcards have shown to improve health literacy and provide individuals with the capacity to act on health information that they may know of but find difficult to process or apply to their daily lives. (2)

Health information has historically been presented such that it is not accessible to most Americans (3). Nearly a third of Americans have low general health literacy (4). These statistics worsen globally. We as a group are motivated to study how masking preferences change based on subconscious visual cues. Unlike in the study quoted above, our participants will not be aware of the true purpose of the study and will see the N95 informational cue as an ad - they are free to make whatever they wish of the cue. Subsequently, on a survey they will be asked their opinion on which category of masks they believe is best at preventing harmful effects of COVID-19 and in another question, their current mask preference. These questions will be interspersed with 10 additional questions on health/wellness behaviors. The survey overall will be presented as an academic study to understand trends in consumer health.

Visuals can stimulate critical thinking and draw more attention than text-only messages. (5) Our goal is to present visual cues in the form of subliminal messages or nudges. Public acceptance of nudges is a well-researched topic (6) and the subliminal presentation of the stimuli is based on ensuring that the stimulus is registered by the appropriate sensory system and activates its corresponding representation, but with minimal activation so that the stimulus does not reach consciousness (7)

In this study, we also consider whether unconscious perceptual processing influences decision-making. We specifically explore the influence of demographic features and personality traits that are related to unconscious processing, namely, the size of the individual’s household, their age as well as the gender of the individual.

Method

Participants:

Since the masking preferences and awareness about covid-19 protection impacts each individual, we decided to collect data through our online Qualtrics survey, across a variety of dimensions and involve individuals from different location, gender, age and ethnicity to name a few. Our participants include people from BU as well as people in our circles which brought good diversity in our data collection. Our survey respondents list included close to 70 individuals, and we received the survey responses from 60 of them.

We understand that there can be some bias in this sample collection since the 60 respondents are not representative of the whole population, but the diverse characteristics of participants helped us to statistically estimate the impact of visual cues on people’s masking preferences.

Treatment:

We included an image displaying benefits of using N-95 masks over other types of masks in the treatment group survey (8). This was the first thing that respondents would see when they received the treatment survey, which was followed up by the 14 generic health questions and 1 outcome question which asked about if respondents think N-95 provides a better protection against covid-19 compared to cotton masks.

Control:

Since the experiment focuses on the impact of visual cues hence keeping that experience consistent across our treatment and control groups, we included a neutral image showing the importance of good health for the people in the control group (9).

Randomization:

Pre-experiment Randomization:

We used Qualtrics' inbuilt randomization for randomly assigning the treatment and control images to our respondents evenly. This option will ensure for every individual seeing the treatment image there will be another individual who sees the control image, hence even randomization, but at any time one individual will see only one kind of image from the two thus making both groups similar with the only difference being the treatment. Other 15 multiple choice questions in our survey were visible to both treatment and control groups in the same sequence with no difference visually (10).

Post-experiment Randomization:

We plan to use blocking after collecting data from all respondents. We will be randomly selecting blocks of respondents based on their characteristics like age, gender, geographical location, size of household, smoking status to name a few. We would consider the time of filling the survey as a randomization variable amongst other characteristics.

Data ingestion

```
dd <- fread("BA830_12_06.csv")
```

Data parsing/cleaning

```
#Removing unwanted columns
colnames(dd)
```

```
##  [1] "StartDate"                   "EndDate"
##  [3] "Status"                      "IPAddress"
##  [5] "Progress"                    "Duration (in seconds)"
##  [7] "Finished"                    "RecordedDate"
##  [9] "ResponseId"                  "RecipientLastName"
## [11] "RecipientFirstName"          "RecipientEmail"
## [13] "ExternalReference"           "LocationLatitude"
## [15] "LocationLongitude"           "DistributionChannel"
## [17] "UserLanguage"                "Q_RelevantIDDuplicate"
## [19] "Q_RelevantIDDuplicateScore"  "Q_RelevantIDFraudScore"
## [21] "Q_RelevantIDLastStartDate"   "Q1"
## [23] "Q2"                          "Q3"
## [25] "Q4"                          "Q5"
## [27] "Q6"                          "Q7"
## [29] "Q8"                          "Q9"
## [31] "Q10"                         "Q11"
## [33] "Q12"                         "Q13"
## [35] "Q14"                         "Q15"
## [37] "DefaultQuestionBlock_DO"
```

```
#Checking for nulls
dd = subset(dd, select = -c(RecipientEmail,       ExternalReference,RecipientLastName,Re
cipientFirstName,Q_RelevantIDDuplicate     ,Q_RelevantIDDuplicateScore,Q_RelevantIDFraud
Score,
Q_RelevantIDLastStartDate,IPAddress,UserLanguage))
```

```
dd <- dd[-c(1, 2),]
dd<-subset(dd, Status!="Survey Preview")
unique(dd$Status)
```

```
## [1] "IP Address"
```

```
#Renaming columns
setnames(dd, old = c('Finished','Q1','Q2','Q3','Q4','Q5','Q6','Q7','Q8','Q9','Q10','Q11'
, 'Q12','Q13','Q14','Q15','DefaultQuestionBlock_DO'), new = c('finished','gender','age',
'ethnicity','region','household_size','annual_health_checkup','mask_type','workout_at_ho
me','languages_spoken','veggies_diet','health_track','resolution','physical_activity','n
95_effective','smoking_habits', 'treatment'))
dd
```

```
##              StartDate            EndDate     Status Progress
##   1: 2021-11-18 18:16:32 2021-11-18 18:16:44 IP Address      100
##   2: 2021-11-18 18:21:44 2021-11-18 18:21:56 IP Address      100
##   3: 2021-11-21 09:53:18 2021-11-21 09:54:43 IP Address      100
##   4: 2021-11-21 10:04:32 2021-11-21 10:05:48 IP Address      100
##   5: 2021-11-21 11:28:05 2021-11-21 11:28:46 IP Address      100
##  ---
## 100: 2021-12-03 17:42:40 2021-12-03 17:44:50 IP Address      100
## 101: 2021-12-03 20:29:56 2021-12-03 20:31:12 IP Address      100
## 102: 2021-11-27 07:42:46 2021-11-27 07:42:56 IP Address       12
## 103: 2021-11-27 09:46:25 2021-11-27 09:47:04 IP Address       35
## 104: 2021-12-04 12:04:14 2021-12-04 12:05:31 IP Address      100
##      Duration (in seconds) finished        RecordedDate        ResponseId
##   1:                    11     True 2021-11-18 18:16:44 R_1g22XHh3gXXIwEb
##   2:                    11     True 2021-11-18 18:21:56 R_1OlwCc9bvL2bveD
##   3:                    84     True 2021-11-21 09:54:43 R_7UqJzymAg70lMKl
##   4:                    75     True 2021-11-21 10:05:48 R_3siw18fo7oU9M97
##   5:                    41     True 2021-11-21 11:28:46 R_2zf1lqnbw3GVptf
##  ---
## 100:                   129     True 2021-12-03 17:44:50 R_2ahn0Zc080jrHf5
## 101:                    75     True 2021-12-03 20:31:12 R_12LL41zMDGz5FZI
## 102:                     9    False 2021-12-04 07:42:57 R_1QGjCTp4XjMg60K
## 103:                    39    False 2021-12-04 09:47:08 R_1PS2T5kT255x3Jx
## 104:                    76     True 2021-12-04 12:05:31 R_2tnjmN5txyOy7lH
##         LocationLatitude     LocationLongitude DistributionChannel gender
##   1:   42.37249755859375      -71.181396484375           anonymous
##   2:  42.346405029296875   -71.097503662109375           anonymous
##   3:     42.356201171875 -71.0631027221679688           anonymous Female
##   4:   42.37249755859375      -71.181396484375           anonymous Female
##   5:   42.37249755859375      -71.181396484375           anonymous Female
##  ---
## 100:   42.47479248046875 -71.44989776611328125           anonymous   Male
## 101: 33.1371002197265625 -96.74880218505859375           anonymous   Male
## 102:                                                     anonymous
## 103:                                                     anonymous   Male
## 104: 42.3513031005859375 -71.13700103759765625           anonymous Female
##           age         ethnicity region household_size annual_health_checkup
##   1:    40-60
##   2: Under 20
##   3:    20-40             Asian     US             No                    No
##   4:    20-40             Asian     US             No                   Yes
##   5:    20-40             Asian Non-US             No                   Yes
##  ---
## 100: Above 60 African American     US             No                   Yes
## 101:    40-60             Asian     US            Yes                   Yes
## 102:
## 103:    20-40             Asian Non-US
## 104:    20-40             Other     US             No                   Yes
##      mask_type workout_at_home languages_spoken veggies_diet health_track
##   1:
##   2:
##   3:                       Yes                1          Yes          Yes
```

```
##   4:                        No                  2          Yes          Yes
##   5:       N95             No                  2          Yes          Yes
##  ---
## 100:  Surgical            Yes                  1          Yes          Yes
## 101:       N95             No      More than 2          Yes           No
## 102:
## 103:
## 104:  Surgical             No      More than 2           No           No
##      resolution physical_activity n95_effective smoking_habits treatment
##   1:
##   2:
##   3:         No                No           Yes   Never smoked        QT
##   4:        Yes               Yes           Yes   Never smoked        QT
##   5:        Yes               Yes           Yes   Never smoked        QT
##  ---
## 100:         No               Yes           Yes   Never smoked        QT
## 101:        Yes         Sometimes           Yes   Never smoked        QT
## 102:                                                                  QT
## 103:                                                                  QT
## 104:        Yes               Yes           Yes Smoke socially        QC
```

```
#counting NA's
dd<-mutate_all(dd,~replace(., . == "", NA))
```

```
dd$na_count <- apply(is.na(dd), 1, sum)
```

Converting outcome and treatment variables to 1s and 0s

```
dd$n95_effective<-ifelse(dd$n95_effective=="Yes",1,0)
dd$treatment<-ifelse(dd$treatment=="QT",1,0)
```

```
dd<-dd[dd$na_count==0 | dd$na_count==17]
unique(dd$na_count)
```

```
## [1]  0 17
```

```
library("VIM")
```

```
## Loading required package: colorspace
```

```
## Loading required package: grid
```

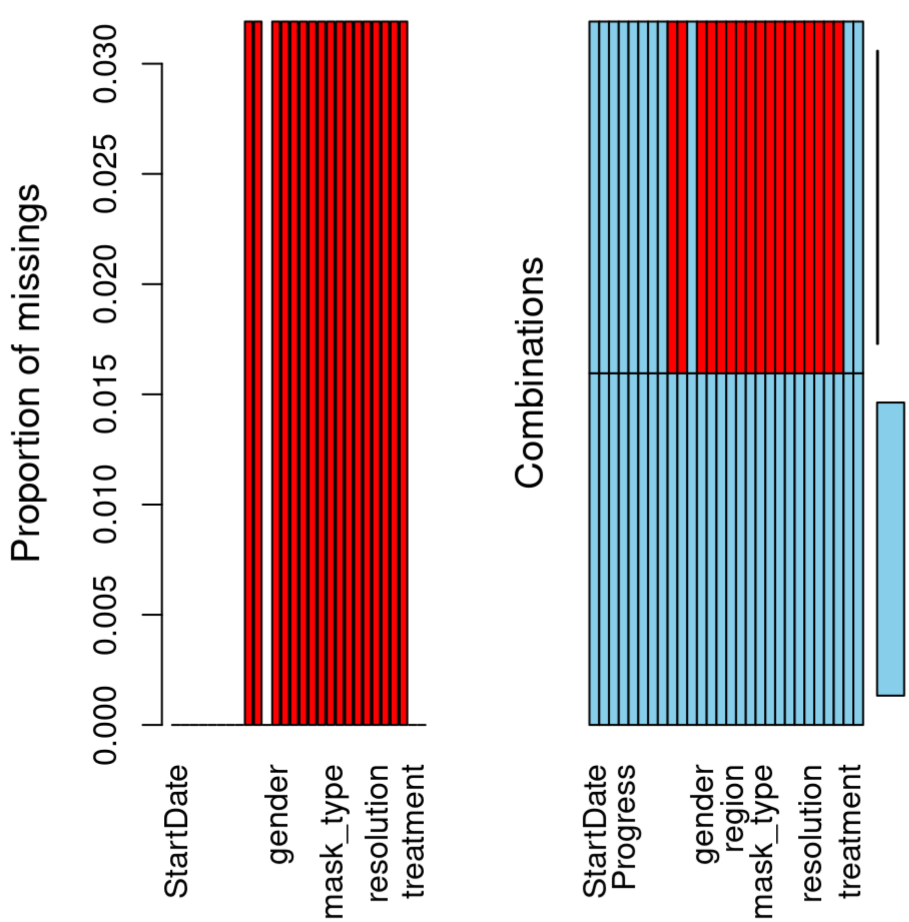
```
## VIM is ready to use.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/i
ssues
```

```
##
## Attaching package: 'VIM'
```

```
## The following object is masked from 'package:datasets':
##
##     sleep
```

```
aggr(dd)
```




```
colnames(dd)
```

```
##  [1] "StartDate"              "EndDate"                "Status"
##  [4] "Progress"               "Duration (in seconds)" "finished"
##  [7] "RecordedDate"           "ResponseId"             "LocationLatitude"
## [10] "LocationLongitude"      "DistributionChannel"    "gender"
## [13] "age"                    "ethnicity"              "region"
## [16] "household_size"         "annual_health_checkup" "mask_type"
## [19] "workout_at_home"        "languages_spoken"       "veggies_diet"
## [22] "health_track"           "resolution"             "physical_activity"
## [25] "n95_effective"          "smoking_habits"         "treatment"
## [28] "na_count"
```

Randomization Checks:

1.Prop test to check for randomization pre-experiment:

```
prop.test(dd[treatment==1,.N],dd[,.N],0.5)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  dd[treatment == 1, .N] out of dd[, .N], null probability 0.5
## X-squared = 0.8617, df = 1, p-value = 0.3533
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.4473567 0.6546328
## sample estimates:
##         p
## 0.5531915
```

Above prop test includes 0.5 hence we can say that randomization was done evenly pre-experiment and survey respondents received both treatment and control images evenly.

2.Randomization check post-experiment:

Randomization check on gender,age and region-Blocking checks:

```
dd$gender1<-ifelse(dd$gender=="Female",1,0)
dd$region1<-ifelse(dd$region=="US",1,0)
dd$age1 <- recode(dd$age,
  "20-40" = 1,
  "40-60" = 2,
  "Above 60" = 3,
  "Under 20" = 0)
dd$ethnicity1 <- recode(dd$ethnicity,
  "Asian" = 0,
  "American" = 1,
  "European" = 2,
  "African American" = 3,
  "Hispanic" = 4,
  "Other" = 5)
```

```
## Warning: Unreplaced values treated as NA as .x is not compatible. Please specify
## replacements exhaustively or supply .default
```

```
random_check_gender1 <- feols(gender1 ~ treatment,data=dd,se="hetero")
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
random_check_age <- feols(age1 ~ treatment,data=dd,se="hetero")
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
random_check_region <- feols(region1 ~ treatment,data=dd,se="hetero")
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3).
```

```
random_check_ethnicity <- feols(ethnicity1 ~ treatment,data=dd,se="hetero")
```

```
## NOTE: 5 observations removed because of NA values (LHS: 5).
```

```
etable(random_check_gender1,random_check_age,random_check_region,random_check_ethnicity)
```

```
##                   random_check_gen..  random_check_age random_check_reg..
## Dependent Var.:              gender1              age1            region1
##
## (Intercept)      0.5122*** (0.0789) 1.244*** (0.0973) 0.5610*** (0.0784)
## treatment           0.0678 (0.1059)   0.0561 (0.1336)   -0.0010 (0.1057)
## _______________  __________________ _________________ __________________
## S.E. type        Heteroskedas.-rob. Heteroskeda.-rob. Heteroskedas.-rob.
## Observations                     91                91                 91
## R2                          0.00460           0.00196            9.56e-7
## Adj. R2                    -0.00659          -0.00925           -0.01124
##                  random_check_et..
## Dependent Var.:        ethnicity1
##
## (Intercept)      0.7000** (0.2559)
## treatment        -0.4755. (0.2817)
## _______________  _________________
## S.E. type        Heteroskeda.-rob.
## Observations                    89
## R2                         0.03559
## Adj. R2                    0.02451
```

From above post- experiment randomization check , we can see that there is statistically significant(at 90%) difference between the ethnicity of the treatment and control group. We do recognize an imbalance in our data in terms of ethnicity and so we tried to circulate the survey more broadly.

Treatment effect analysis

In this experiment, we have an established treatment and placebo group and we are aware of the complier rates (96% and 98% in the T and C respectively). We can directly calculate the CACE by comparing the ITTs of the two groups. # Complier rates

```
cr_t<-dd[finished == "True" & treatment ==1 ,.N]/dd[treatment ==1 ,.N]
cr_c<-dd[finished == "True" & treatment ==0 ,.N]/dd[treatment ==0 ,.N]
```

CACE calculation

The effect of the treatment i.e. the n95 focused image before the survey outcome which is the answer of the question “Do you think that n95 masks are more effective than cotton masks in preventing COVID -19?” seems to statistically and economically insignificant with a very high standard error. A whopping 82% of our respondents answered yes to the outcome question irrespective of being treated. The a greater percentage(~2%) of the treatment group did however yes to the outcome question.

```
this_reg <- feols(n95_effective~treatment, data = dd[finished == "True"], se = "hetero")
R2reg <- data.frame(summary(this_reg)$r.squared,summary(this_reg)$adj.r.squared)
colnames(R2reg)[0]<- 'R Squared'
colnames(R2reg)[0]<- 'Adjusted R Squared'
this_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c') %>%
kable_classic_2(full_width=F)
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.829	0.059	13.956	0.000
treatment	0.011	0.079	0.135	0.893

Checking the ratio of Yes and No responses for our outcome question

```
dd[finished == "True" & n95_effective == 1,.N]/dd[finished == "True",.N]
```

```
## [1] 0.8351648
```

```
dd[finished == "True" & n95_effective == 1,.N,treatment][1,2]/dd[finished == "True",.N,t
reatment][1,2]
```

```
##       N
## 1: 0.84
```

```
dd[finished == "True" & n95_effective == 1,.N,treatment][2,2]/dd[finished == "True",.N,t
reatment][2,2]
```

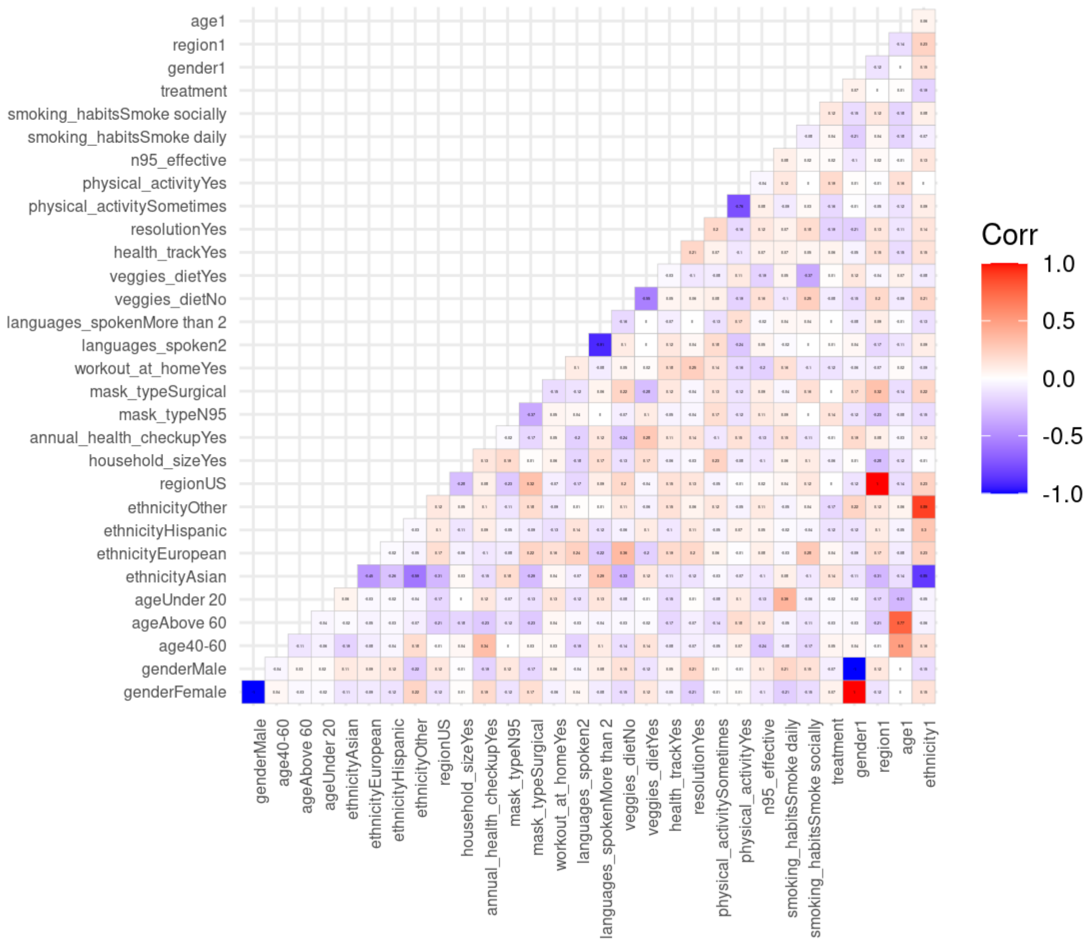
```
##            N
## 1: 0.8292683
```

Covariate analysis

As a first step, we observe correlation within the covariates and the outcome to identify broad trends. Household size, workout at home, and no veggies in the diet are negatively correlated with the outcome whereas health tracking is positively correlated. Since our survey collects demographic and behavioral data other than the outcome, the experiment should have no impact on them and hence, we do not have bad covariates.

```
dd_cov <- dd[finished == "True",!c(1,2,3,4,5,6,7,8,9,10,11)]
library(ggcorrplot)
model.matrix(~0+., data=dd_cov) %>%
  cor(use="pairwise.complete.obs") %>%
ggcorrplot(show.diag = F, type="lower", lab=TRUE, lab_size=0.5, tl.cex = 6, tl.srt = 90)
```

```
## Warning in cor(., use = "pairwise.complete.obs"): the standard deviation is zero
```



Adding in covariates with higher correlation with the outcome did not improve precision of our CACE prediction. The intercept value however that gone up to 96% pointing to that the broader population beliefs in n95 effectiveness.

```
this_reg_cov <- feols(n95_effective~treatment+age+household_size+annual_health_checkup+v
eggies_diet+workout_at_home, data = dd[finished == "True"], se = "hetero")
R2reg_cov <- data.frame(summary(this_reg_cov)$r.squared,summary(this_reg_cov)$adj.r.squa
red)
colnames(R2reg_cov)[0]<- 'R Squared'
colnames(R2reg_cov)[0]<- 'Adjusted R Squared'
this_reg_cov %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Sta
t",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c') %>%
kable_classic_2(full_width=F)
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.957	0.107	8.920	0.000
treatment	0.009	0.076	0.121	0.904
age40-60	-0.235	0.149	-1.573	0.120
ageAbove 60	0.168	0.078	2.137	0.036
ageUnder 20	-0.309	0.428	-0.722	0.472
household_sizeYes	-0.027	0.080	-0.332	0.741

Predictor	Coefficient	SE	T-Stat	P-Value
annual_health_checkupYes	0.042	0.078	0.535	0.594
veggies_dietNo	0.071	0.090	0.789	0.432
veggies_dietYes	-0.087	0.090	-0.968	0.336
workout_at_homeYes	-0.138	0.072	-1.907	0.060

Heterogenous Effects:

CATE on variables that are highly correlated with our outcome variable : n95_effective. After conducting the interactions between treatment and other covariates we could see statistically significant impact only due to "Age". Hence we decide to include that in our interaction term for calculating heterogenous effect and CATE.

```
reg_age <- feols(n95_effective ~ treatment*age ,data=dd , se="hetero")
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3, RHS: 3).
```

```
R2reg_age <- data.frame(summary(reg_age)$r.squared,summary(reg_age)$adj.r.squared)
colnames(R2reg_age)[0]<- 'R Squared'
colnames(R2reg_age)[0]<- 'Adjusted R Squared'
reg_age %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c') %>%
kable_classic_2(full_width=F)
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.844	0.067	12.554	0.000
treatment	0.048	0.086	0.561	0.577
age40-60	-0.244	0.239	-1.020	0.311
ageAbove 60	0.156	0.067	2.325	0.023
ageUnder 20	0.156	0.067	2.325	0.023
treatment:age40-60	-0.023	0.304	-0.076	0.939
treatment:ageAbove 60	-0.048	0.086	-0.561	0.577
treatment:ageUnder 20	-1.048	0.086	-12.206	0.000

The current mask type (if n95) has a large bearing on the masking beliefs but the findings are statistically insignificant.

```
reg_u <- feols(n95_effective ~ treatment*mask_type ,data=dd , se="hetero")
```

```
## NOTE: 3 observations removed because of NA values (LHS: 3, RHS: 3).
```

```
R3reg_masks <- data.frame(summary(reg_u)$r.squared,summary(reg_u)$adj.r.squared)
colnames(R3reg_masks)[0]<- 'R Squared'
colnames(R3reg_masks)[0]<- 'Adjusted R Squared'
reg_u %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat",
"P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c') %>%
kable_classic_2(full_width=F)
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.800	0.093	8.644	0.000
treatment	-0.078	0.143	-0.543	0.588
mask_typeN95	-0.050	0.242	-0.206	0.837
mask_typeSurgical	0.082	0.123	0.670	0.505
treatment:mask_typeN95	0.328	0.266	1.233	0.221
treatment:mask_typeSurgical	0.059	0.181	0.326	0.745

Potential risks and improvement opportunities

- Selection bias Our survey data contains inherent selection bias - A large majority of our respondents(>80%) identify as Asian, and a most of our data is concentrated in the ~20-60 age bucket. This could potentially skew our results because of a non-representative sample.
- Randomization Because of the selection bias in data, the treatment and control group differ in ethnicity at 90% significance. This could skew results because of differences in health access across ethnicities.
- Sample size The cleaned dataset with complete data contains the information of 94 individuals. This is quite a small sample size to measure a sizable effect. The power of this experiment is 0.08 to detect an effect size of 0.1.

```
pwr.t2n.test(n1 = dd[treatment == 1,.N], n2 =dd[treatment == 0,.N],d=0.1,sig.level = 0.0
5)
```

```
##
##      t test power calculation
##
##              n1 = 52
##              n2 = 42
##               d = 0.1
##       sig.level = 0.05
##           power = 0.07644803
##     alternative = two.sided
```

- Weak visual cue The image used as a visual cue for n95 mask preference in the experiment may not be strong enough to command a response and we could potentially experiment with more treatment images

Key inferences and conclusions

- Experiment design and response Our experiment included an easy-to-fill survey from a clickable link. The survey has less than 250 words and takes under 5 minutes to fill. This user friendly format ensured a high complier rate in both the treatment and placebo group
- Treatment effect The effect of the treatment i.e. the n95 focused image before the survey outcome which is the answer of the question “Do you think that n95 masks are more effective than cotton masks in preventing COVID -19?” is statistically and economically insignificant with a very high standard error.

A whopping 82% of our respondents answered yes to the outcome question irrespective of being treated.

The a greater percentage(~2%) of the treatment group did however yes to the outcome question.

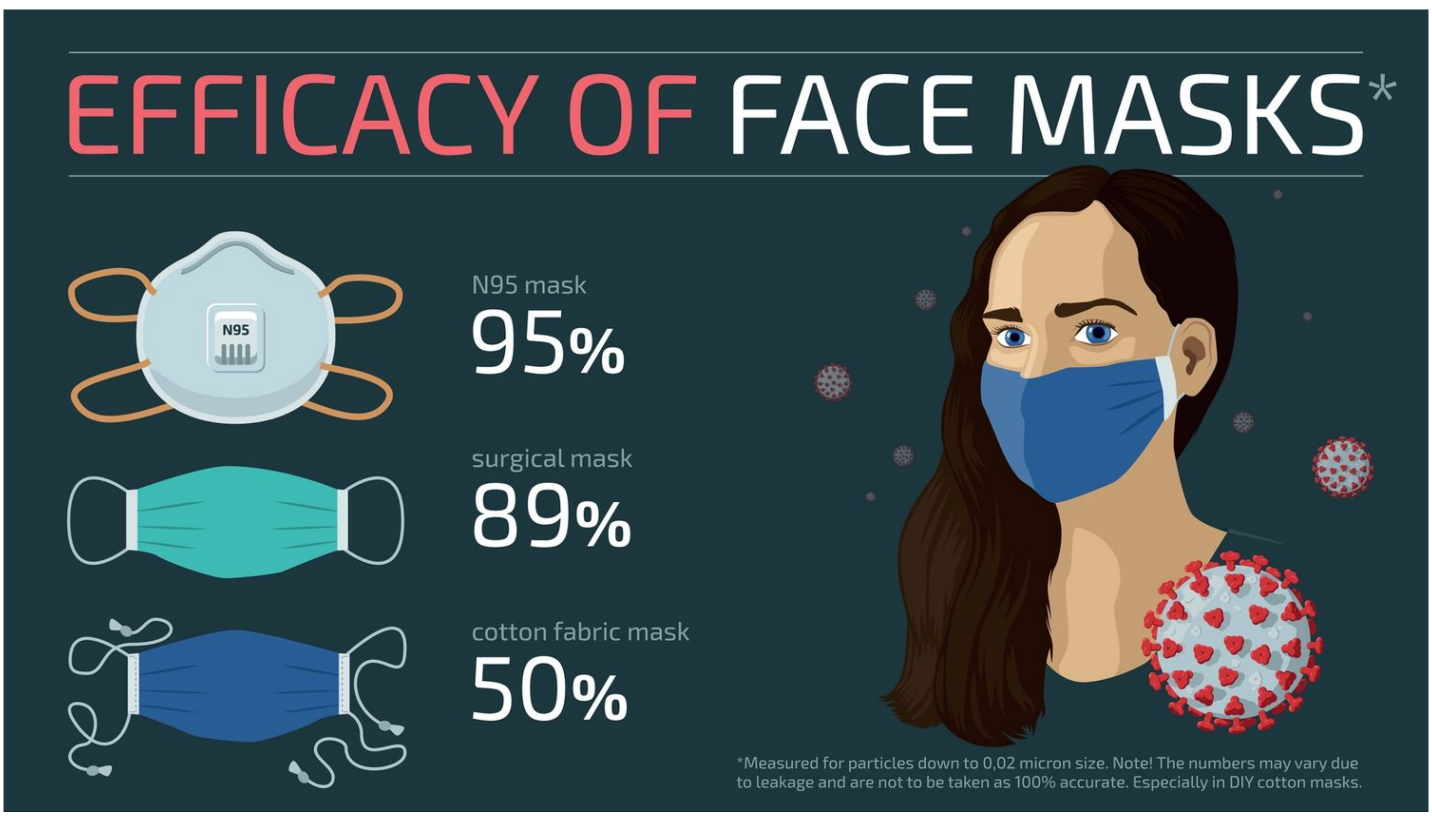
- Impact of covariates Age, Household size, workout at home, and no veggies in the diet are negatively correlated with the outcome whereas health tracking is positively correlated.
- Conditional effects Current masking behaviors seem to have a large economic impact on masking beliefs, i.e. current n95 users showed a large preference to saying yes to the outcome question. These results are however, not statistically significant.

Overall, we have found some promising results on the impact of visual cues on masking preferences. We have been able to combine this with interesting lifestyle data and we look forward to further building this analysis.

References

(1)Imram, N. (1999), "The role of visual cues in consumer perception and acceptance of a food product", Nutrition & Food Science, Vol. 99 No. 5, pp. 224-230 https://doi.org/10.1108/00346659910277650 (https://doi.org/10.1108/00346659910277650)

2. (2021). "Visual Narratives About COVID-19 Improve Message Accessibility, Self-Efficacy, and Health Precautions" https://doi.org/10.3389/fcomm.2021.712658 (https://doi.org/10.3389/fcomm.2021.712658)
3. "Health Literacy" by CDC (2021)
4. Paasche-Orlow et al., 2005
5. Cvijikj and Michahelles, 2013; Kim et al., 2015; Lazard and Atkinson, 2015
6. Sunstein CR. People Prefer System 2 Nudges (Kind of). Duke LJ. 2016; 66:121–168
7. Smith and McCulloch, 2012
8. Treatment group image



Treatment Image

9. Control group image

Control Image

10. Survey questions interface common for both treatment and control groups:

Boston University

What gender do you identify with?

- Male
- Female
- Other

What is your age?

- Under 20
- 20-40
- 40-60
- Above 60

What is your ethnicity?

- Asian
- American
- African American
- European
- Hispanic

Do you live in a household with more than 3 people?

- Yes
- No

Do you go for an annual general health-checkup ?

- Yes
- No

What is your current choice of go-to masks ?

- Cloth
- N95
- Surgical

Do you prefer working out at home over the gym?

- Yes
- No

How many languages do you speak?

- ○ 1
- ○ 2
- ○ More than 2

Do you think your diet contains enough fruits and vegetables?

- ○ Yes
- ○ No
- ○ May be

Do you use a mobile app for tracking your daily foot count or monitoring health activity?

- ○ Yes
- ○ No

Do your new year resolutions include health/physical activity/weight loss-related goals?

- ○ Yes
- ○ No

Do you think you get 20 mins of physical activity(walking, running, etc.) 4-5 times a week ?

- Yes
- No
- Sometimes

Do you think that n95 masks are more effective than cotton masks in preventing COVID-19?

- Yes
- No

Select what describes your smoking habits?

- Never smoked
- Smoke socially
- Smoke daily

→

Boston University

We thank you for your time spent taking this survey.
Your response has been recorded.

11. CDC 2020b, 2020c; Edelstein and Ramakrishnan 2020; WHO 2020

12. John T. Brooks, MD1; Jay C. Butler, MD - Effectiveness of Mask Wearing to Control Community Spread of SARS-CoV-2