```r
#load package
library(dplyr)
library(tidyr)
library(gtsummary)
library(labelled)
library(MASS)
library(stats)
library(ggplot2)
#library(effects)
library(car)
library(tibble)
```

```r
source("./function.R")
```

```
## Loading required package: tidyverse
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.5
## v lubridate 1.9.3     v stringr   1.5.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x car::recode()   masks dplyr::recode()
## x MASS::select()  masks gtsummary::select(), dplyr::select()
## x purrr::some()   masks car::some()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## Loading required package: rcompanion
```

```r
#Load preprocessed data
#data_filter <- readRDS(file = "../output/data_preprocessed.rds")
#data_filter <- readRDS(file = "../output/data_preprocessed_edited.rds") # merge the myanmar and the he
data_filter <- readRDS(file = "../output/data_preprocessed_mergeethnic.rds") # merge the myanmar
```

```r
## define custom test
fisher.test.simulate.p.values <- function(data, variable, by, ...) {
  result <- list()
  test_results <- stats::fisher.test(data[[variable]], data[[by]], simulate.p.value = TRUE)
  result$p <- test_results$p.value
  result$test <- test_results$method
  result
}
```

```r
##table 1 and table 2

#temporary convert Phq91-9 to factor
data_filter_phq9asfactor <- data_filter
data_filter_phq9asfactor$phq9_1 <- factor(data_filter_phq9asfactor$phq9_1, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_2 <- factor(data_filter_phq9asfactor$phq9_2, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_3 <- factor(data_filter_phq9asfactor$phq9_3, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_4 <- factor(data_filter_phq9asfactor$phq9_4, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_5 <- factor(data_filter_phq9asfactor$phq9_5, levels = c("0","1","2","3"))
```

```r
data_filter_phq9asfactor$phq9_6 <- factor(data_filter_phq9asfactor$phq9_6, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_7 <- factor(data_filter_phq9asfactor$phq9_7, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_8 <- factor(data_filter_phq9asfactor$phq9_8, levels = c("0","1","2","3"))
data_filter_phq9asfactor$phq9_9 <- factor(data_filter_phq9asfactor$phq9_9, levels = c("0","1","2","3"))
```

```r
table1 <-
  data_filter_phq9asfactor %>%
  tbl_summary()

table2 <-
  data_filter_phq9asfactor %>%
  tbl_summary(by = phq_9_cat) %>%
  add_p(
  test = list(all_categorical() ~ "fisher.test.simulate.p.values")  # this applies the custom test to a
)  %>%
  add_overall()
```

```r
table1
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 272**
Age	73 (67, 78)
Sex	
male	78 (29%)
female	194 (71%)
Weight of Patient (kgs)	58 (51, 66)
Height of Patient (cms)	156 (151, 162)
BMI	23.5 (21.3, 25.7)
Ethnic	
Thai	263 (97%)
Chinese	9 (3.3%)
Marital status	
single	50 (18%)
married	172 (63%)
divorced	10 (3.7%)
widow	40 (15%)
Address	
Bangkok	182 (67%)
others	90 (33%)
Education	
not educate	3 (1.1%)
elementary	36 (13%)
high school	42 (15%)
college degree	191 (70%)
above college degree	0 (0%)
Employment	
unemployed	93 (34%)
part-time job	23 (8.5%)

Characteristic	**N = 272**
full-time job	24 (8.8%)
retired	132 (49%)
income	
10,000 or less	52 (19%)
10,001 - 20,000	50 (18%)
20,001 - 30,000	44 (16%)
30,001 or more	66 (24%)
unknown	60 (22%)
Income Loss from COVID-19	
Same	221 (81%)
Less than 50% loss	18 (6.6%)
Over 50% loss	20 (7.4%)
No income	13 (4.8%)
Ambulation	
Normal	260 (96%)
Gait aid	12 (4.4%)
Bedbound	0 (0%)
Hearing	
Normal	255 (94%)
Hearing aid	2 (0.7%)
Hearing impairment	15 (5.5%)
Visual	
Normal	172 (63%)
Glasses	100 (37%)
Vision loss	0 (0%)
Smoking	
Never smoking	234 (86%)
Current smoking	6 (2.2%)
Past smoking	32 (12%)
Alcohol Drinking	
Never drinking	247 (91%)
Social drinking	22 (8.1%)
Regular drinking	3 (1.1%)
Dementia diagnosis	
No	255 (94%)
Yes	10 (3.7%)
Not sure	7 (2.6%)
Self Percept Cognition	
Normal	115 (42%)
Minor cognitive problem	155 (57%)
Major cognitive problem	2 (0.7%)
Number of Hospitalization	
0	230 (85%)
1	35 (13%)
2	5 (1.8%)
3	2 (0.7%)
Self Percept Health	
Worst	1 (0.4%)
Bad	6 (2.2%)
Average	111 (41%)
Good	131 (48%)
Best	23 (8.5%)

Characteristic	N = 272
neuro	
None	238 (88%)
Neurological disease	34 (13%)
cvs	
None	53 (19%)
Cardiovascular disease	219 (81%)
respi	
None	251 (92%)
Respiratory disease	21 (7.7%)
gi	
None	216 (79%)
Gastrointestinal disease	56 (21%)
renal	
None	250 (92%)
Renal disease	22 (8.1%)
endo	
None	211 (78%)
Endocrine disease	61 (22%)
msk	
None	172 (63%)
MSK disease	100 (37%)
cancer	
None	249 (92%)
Cancer	23 (8.5%)
allergy	
None	197 (72%)
Allergy	75 (28%)
psychi	
None	257 (94%)
Psych disease	15 (5.5%)
phq9_1	
0	186 (68%)
1	78 (29%)
2	6 (2.2%)
3	2 (0.7%)
phq9_2	
0	212 (78%)
1	56 (21%)
2	4 (1.5%)
3	0 (0%)
phq9_3	
0	171 (63%)
1	65 (24%)
2	23 (8.5%)
3	13 (4.8%)
phq9_4	
0	208 (76%)
1	47 (17%)
2	13 (4.8%)
3	4 (1.5%)
phq9_5	
0	222 (82%)

Characteristic	N = 272
1	37 (14%)
2	8 (2.9%)
3	5 (1.8%)
phq9_6	
0	254 (93%)
1	15 (5.5%)
2	3 (1.1%)
3	0 (0%)
phq9_7	
0	250 (92%)
1	20 (7.4%)
2	1 (0.4%)
3	1 (0.4%)
phq9_8	
0	254 (93%)
1	15 (5.5%)
2	2 (0.7%)
3	1 (0.4%)
phq9_9	
0	268 (99%)
1	4 (1.5%)
2	0 (0%)
3	0 (0%)
PHQ-9 score	1 (0, 3)
PHQ-9 Interpretation	
normal	233 (86%)
mild depression	33 (12%)
moderate depression	6 (2.2%)

```
table2
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**Overall**, N = 272	**normal**, N = 233	**mild depression**, N = 33	**moderate depression**, N = 6	**p-value**
Age	73 (67, 78)	72 (67, 77)	74 (67, 80)	80 (76, 85)	0.020
Sex					>0.9
male	78 (29%)	67 (29%)	9 (27%)	2 (33%)	
female	194 (71%)	166 (71%)	24 (73%)	4 (67%)	
Weight of Patient (kgs)	58 (51, 66)	58 (51, 66)	60 (52, 70)	50 (48, 58)	0.3
Height of Patient (cms)	156 (151, 162)	156 (151, 163)	155 (150, 160)	159 (152, 164)	0.3
BMI	23.5 (21.3, 25.7)	23.4 (21.2, 25.6)	24.6 (22.9, 27.3)	20.9 (19.9, 22.0)	0.017
Ethnic					0.3
Thai	263 (97%)	226 (97%)	32 (97%)	5 (83%)	
Chinese	9 (3.3%)	7 (3.0%)	1 (3.0%)	1 (17%)	

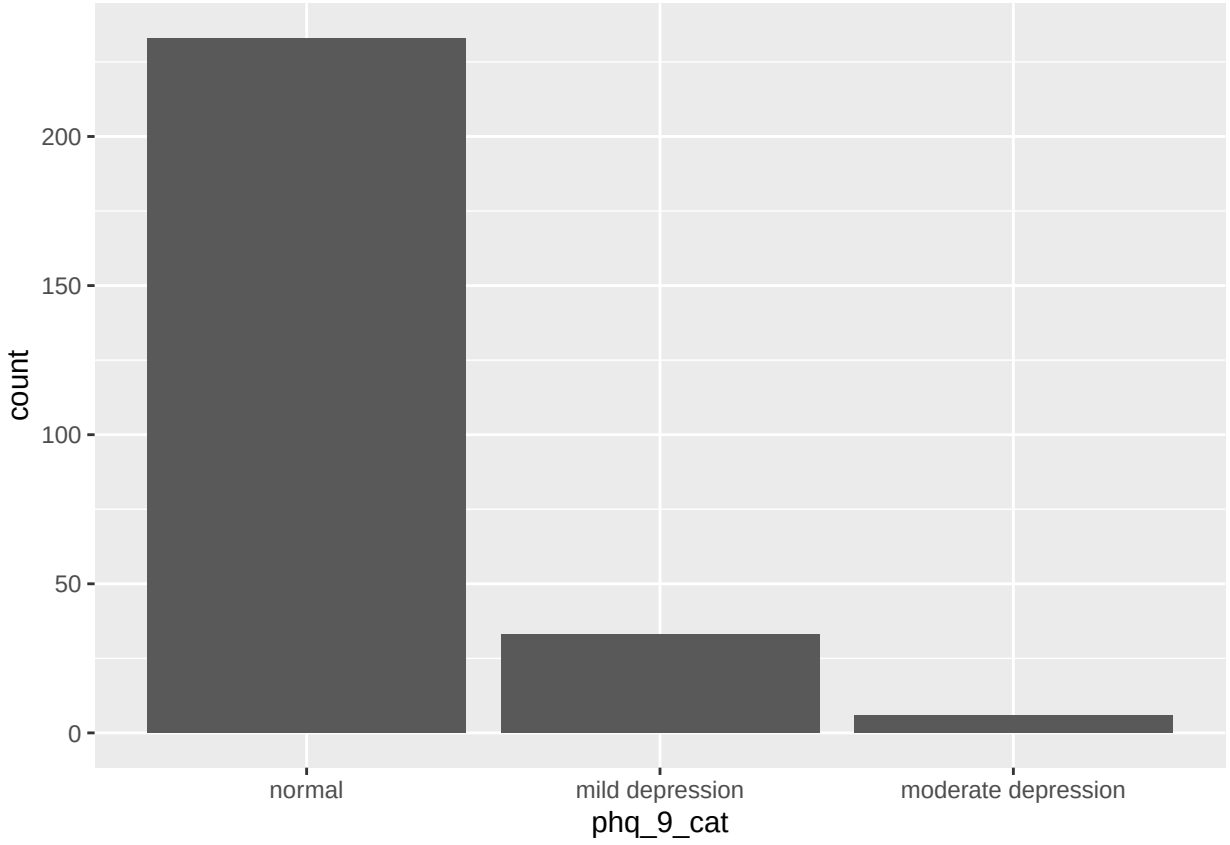
Characteristic	**Overall**, N = 272	**normal**, N = 233	**mild depression**, N = 33	**moderate depression**, N = 6	**p-value**
Marital status					0.5
single	50 (18%)	45 (19%)	5 (15%)	0 (0%)	
married	172 (63%)	149 (64%)	19 (58%)	4 (67%)	
divorced	10 (3.7%)	8 (3.4%)	2 (6.1%)	0 (0%)	
widow	40 (15%)	31 (13%)	7 (21%)	2 (33%)	
Address					0.2
Bangkok	182 (67%)	152 (65%)	24 (73%)	6 (100%)	
others	90 (33%)	81 (35%)	9 (27%)	0 (0%)	
Education					0.13
not educate	3 (1.1%)	2 (0.9%)	1 (3.0%)	0 (0%)	
elementary	36 (13%)	29 (12%)	6 (18%)	1 (17%)	
high school	42 (15%)	32 (14%)	9 (27%)	1 (17%)	
college degree	191 (70%)	170 (73%)	17 (52%)	4 (67%)	
above college degree	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Employment					0.3
unemployed	93 (34%)	75 (32%)	17 (52%)	1 (17%)	
part-time job	23 (8.5%)	19 (8.2%)	3 (9.1%)	1 (17%)	
full-time job	24 (8.8%)	22 (9.4%)	2 (6.1%)	0 (0%)	
retired	132 (49%)	117 (50%)	11 (33%)	4 (67%)	
income					0.078
10,000 or less	52 (19%)	40 (17%)	11 (33%)	1 (17%)	
10,001 - 20,000	50 (18%)	43 (18%)	4 (12%)	3 (50%)	
20,001 - 30,000	44 (16%)	39 (17%)	5 (15%)	0 (0%)	
30,001 or more	66 (24%)	62 (27%)	4 (12%)	0 (0%)	
unknown	60 (22%)	49 (21%)	9 (27%)	2 (33%)	
Income Loss from COVID-19					0.2
Same	221 (81%)	194 (83%)	22 (67%)	5 (83%)	
Less than 50% loss	18 (6.6%)	14 (6.0%)	4 (12%)	0 (0%)	
Over 50% loss	20 (7.4%)	15 (6.4%)	4 (12%)	1 (17%)	
No income	13 (4.8%)	10 (4.3%)	3 (9.1%)	0 (0%)	
Ambulation					0.4
Normal	260 (96%)	224 (96%)	30 (91%)	6 (100%)	
Gait aid	12 (4.4%)	9 (3.9%)	3 (9.1%)	0 (0%)	
Bedbound	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Hearing					0.2
Normal	255 (94%)	221 (95%)	29 (88%)	5 (83%)	
Hearing aid	2 (0.7%)	2 (0.9%)	0 (0%)	0 (0%)	
Hearing impairment	15 (5.5%)	10 (4.3%)	4 (12%)	1 (17%)	
Visual					0.5
Normal	172 (63%)	144 (62%)	24 (73%)	4 (67%)	
Glasses	100 (37%)	89 (38%)	9 (27%)	2 (33%)	
Vision loss	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Smoking					0.5
Never smoking	234 (86%)	202 (87%)	26 (79%)	6 (100%)	
Current smoking	6 (2.2%)	5 (2.1%)	1 (3.0%)	0 (0%)	
Past smoking	32 (12%)	26 (11%)	6 (18%)	0 (0%)	
Alcohol Drinking					0.5

Characteristic	**Overall**, N = 272	**normal**, N = 233	**mild depression**, N = 33	**moderate depression**, N = 6	**p-value**
Never drinking	247 (91%)	212 (91%)	29 (88%)	6 (100%)	
Social drinking	22 (8.1%)	19 (8.2%)	3 (9.1%)	0 (0%)	
Regular drinking	3 (1.1%)	2 (0.9%)	1 (3.0%)	0 (0%)	
Dementia diagnosis					0.041
No	255 (94%)	222 (95%)	28 (85%)	5 (83%)	
Yes	10 (3.7%)	7 (3.0%)	2 (6.1%)	1 (17%)	
Not sure	7 (2.6%)	4 (1.7%)	3 (9.1%)	0 (0%)	
Self Percept Cognition					0.002
Normal	115 (42%)	107 (46%)	7 (21%)	1 (17%)	
Minor cognitive problem	155 (57%)	125 (54%)	26 (79%)	4 (67%)	
Major cognitive problem	2 (0.7%)	1 (0.4%)	0 (0%)	1 (17%)	
Number of Hospitalization					0.002
0	230 (85%)	204 (88%)	22 (67%)	4 (67%)	
1	35 (13%)	24 (10%)	9 (27%)	2 (33%)	
2	5 (1.8%)	5 (2.1%)	0 (0%)	0 (0%)	
3	2 (0.7%)	0 (0%)	2 (6.1%)	0 (0%)	
Self Percept Health					<0.001
Worst	1 (0.4%)	0 (0%)	1 (3.0%)	0 (0%)	
Bad	6 (2.2%)	2 (0.9%)	1 (3.0%)	3 (50%)	
Average	111 (41%)	88 (38%)	22 (67%)	1 (17%)	
Good	131 (48%)	121 (52%)	9 (27%)	1 (17%)	
Best	23 (8.5%)	22 (9.4%)	0 (0%)	1 (17%)	
neuro					0.053
None	238 (88%)	208 (89%)	26 (79%)	4 (67%)	
Neurological disease	34 (13%)	25 (11%)	7 (21%)	2 (33%)	
cvs					0.6
None	53 (19%)	48 (21%)	4 (12%)	1 (17%)	
Cardiovascular disease	219 (81%)	185 (79%)	29 (88%)	5 (83%)	
respi					0.007
None	251 (92%)	218 (94%)	30 (91%)	3 (50%)	
Respiratory disease	21 (7.7%)	15 (6.4%)	3 (9.1%)	3 (50%)	
gi					0.003
None	216 (79%)	189 (81%)	26 (79%)	1 (17%)	
Gastrointestinal disease	56 (21%)	44 (19%)	7 (21%)	5 (83%)	
renal					0.018
None	250 (92%)	218 (94%)	28 (85%)	4 (67%)	
Renal disease	22 (8.1%)	15 (6.4%)	5 (15%)	2 (33%)	
endo					0.3
None	211 (78%)	184 (79%)	22 (67%)	5 (83%)	
Endocrine disease	61 (22%)	49 (21%)	11 (33%)	1 (17%)	

Characteristic	**Overall**, N = 272	**normal**, N = 233	**mild depression**, N = 33	**moderate depression**, N = 6	**p-value**
msk					0.014
None	172 (63%)	155 (67%)	15 (45%)	2 (33%)	
MSK disease	100 (37%)	78 (33%)	18 (55%)	4 (67%)	
cancer					0.7
None	249 (92%)	214 (92%)	29 (88%)	6 (100%)	
Cancer	23 (8.5%)	19 (8.2%)	4 (12%)	0 (0%)	
allergy					0.12
None	197 (72%)	171 (73%)	24 (73%)	2 (33%)	
Allergy	75 (28%)	62 (27%)	9 (27%)	4 (67%)	
psychi					0.006
None	257 (94%)	224 (96%)	29 (88%)	4 (67%)	
Psych disease	15 (5.5%)	9 (3.9%)	4 (12%)	2 (33%)	
phq9_1					<0.001
0	186 (68%)	178 (76%)	6 (18%)	2 (33%)	
1	78 (29%)	52 (22%)	25 (76%)	1 (17%)	
2	6 (2.2%)	3 (1.3%)	1 (3.0%)	2 (33%)	
3	2 (0.7%)	0 (0%)	1 (3.0%)	1 (17%)	
phq9_2					<0.001
0	212 (78%)	194 (83%)	15 (45%)	3 (50%)	
1	56 (21%)	38 (16%)	17 (52%)	1 (17%)	
2	4 (1.5%)	1 (0.4%)	1 (3.0%)	2 (33%)	
3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
phq9_3					<0.001
0	171 (63%)	165 (71%)	6 (18%)	0 (0%)	
1	65 (24%)	50 (21%)	15 (45%)	0 (0%)	
2	23 (8.5%)	15 (6.4%)	6 (18%)	2 (33%)	
3	13 (4.8%)	3 (1.3%)	6 (18%)	4 (67%)	
phq9_4					<0.001
0	208 (76%)	198 (85%)	9 (27%)	1 (17%)	
1	47 (17%)	32 (14%)	14 (42%)	1 (17%)	
2	13 (4.8%)	1 (0.4%)	9 (27%)	3 (50%)	
3	4 (1.5%)	2 (0.9%)	1 (3.0%)	1 (17%)	
phq9_5					<0.001
0	222 (82%)	209 (90%)	11 (33%)	2 (33%)	
1	37 (14%)	21 (9.0%)	14 (42%)	2 (33%)	
2	8 (2.9%)	2 (0.9%)	5 (15%)	1 (17%)	
3	5 (1.8%)	1 (0.4%)	3 (9.1%)	1 (17%)	
phq9_6					<0.001
0	254 (93%)	226 (97%)	25 (76%)	3 (50%)	
1	15 (5.5%)	7 (3.0%)	6 (18%)	2 (33%)	
2	3 (1.1%)	0 (0%)	2 (6.1%)	1 (17%)	
3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
phq9_7					<0.001
0	250 (92%)	226 (97%)	22 (67%)	2 (33%)	
1	20 (7.4%)	7 (3.0%)	11 (33%)	2 (33%)	
2	1 (0.4%)	0 (0%)	0 (0%)	1 (17%)	
3	1 (0.4%)	0 (0%)	0 (0%)	1 (17%)	
phq9_8					<0.001
0	254 (93%)	228 (98%)	24 (73%)	2 (33%)	
1	15 (5.5%)	5 (2.1%)	9 (27%)	1 (17%)	
2	2 (0.7%)	0 (0%)	0 (0%)	2 (33%)	

Characteristic	Overall, N = 272	normal, N = 233	mild depression, N = 33	moderate depression, N = 6	p-value
3	1 (0.4%)	0 (0%)	0 (0%)	1 (17%)	
phq9_9					0.009
0	268 (99%)	232 (100%)	31 (94%)	5 (83%)	
1	4 (1.5%)	1 (0.4%)	2 (6.1%)	1 (17%)	
2	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
3	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
PHQ-9 score	1 (0, 3)	1 (0, 2)	6 (5, 7)	11 (9, 13)	<0.001

```
data_filter %>% ggplot(aes(x = phq_9_cat)) + geom_bar()
```

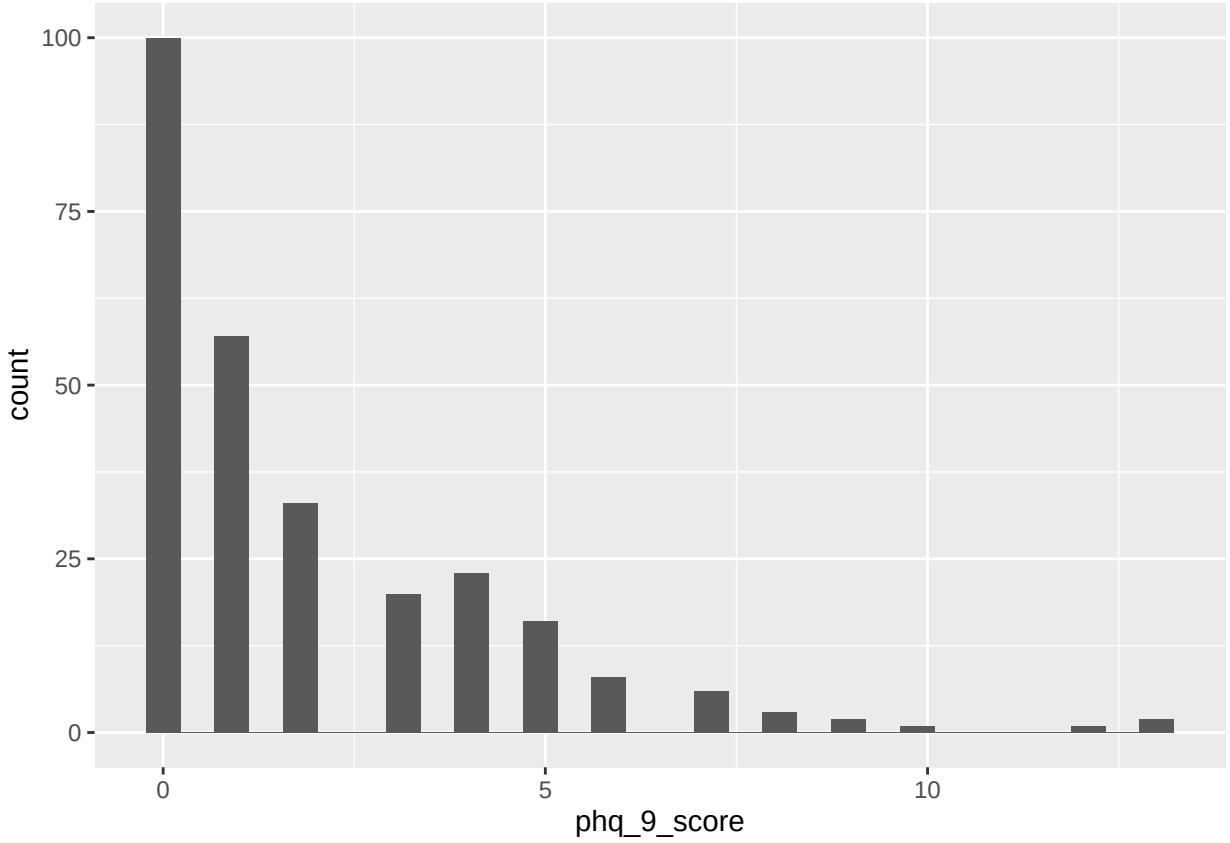



```
data_filter %>% group_by(phq_9_cat) %>%
  summarise(frequency = n())
```

```
## # A tibble: 3 x 2
##   phq_9_cat           frequency
##   <fct>                   <int>
## 1 normal                    233
## 2 mild depression            33
## 3 moderate depression         6
```

```
data_filter %>% ggplot(aes(x = phq_9_score)) + geom_histogram()
```

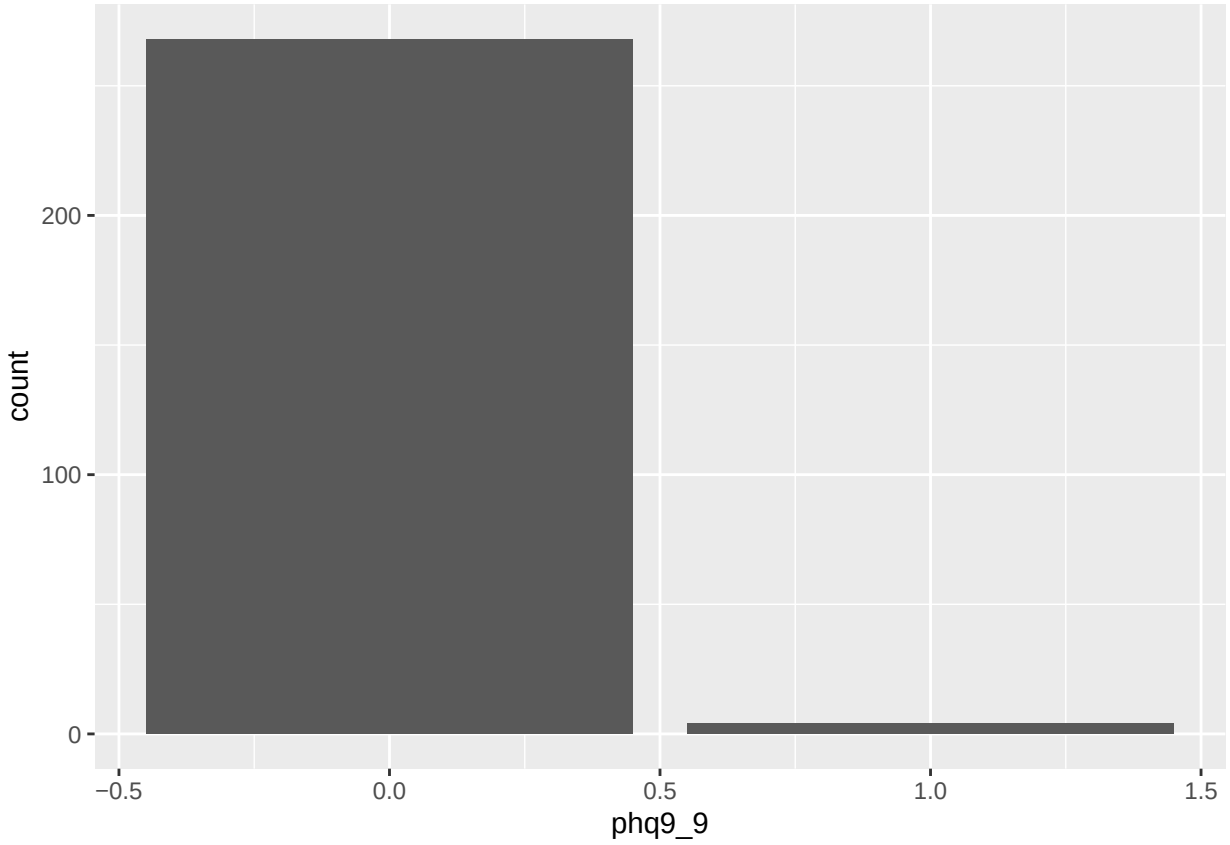
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
data_filter %>% group_by(phq_9_score) %>%
  summarise(frequency = n())
```

```
## # A tibble: 13 x 2
##    phq_9_score frequency
##          <int>     <int>
##  1           0       100
##  2           1        57
##  3           2        33
##  4           3        20
##  5           4        23
##  6           5        16
##  7           6         8
##  8           7         6
##  9           8         3
## 10           9         2
## 11          10         1
## 12          12         1
## 13          13         2
```

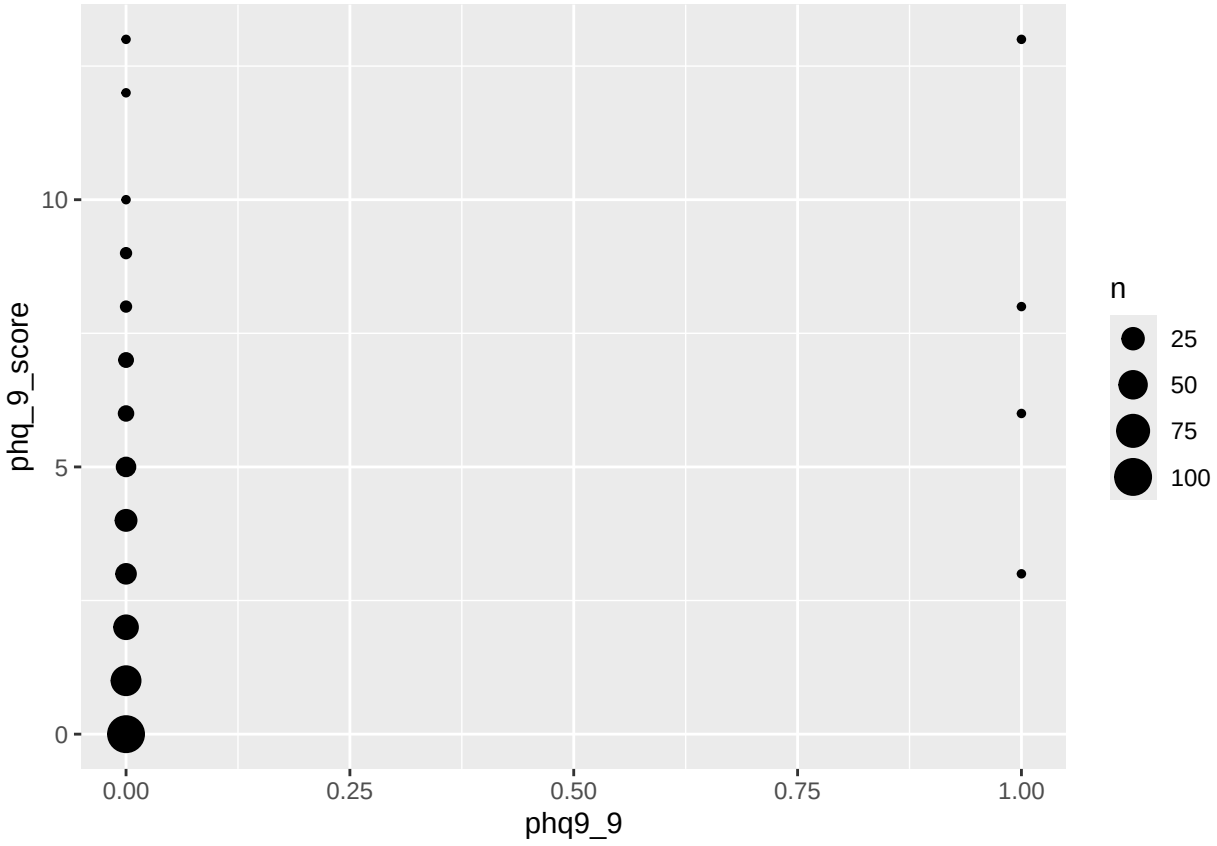
```
data_filter %>% ggplot(aes(x = phq9_9)) + geom_bar()
```




```
data_filter %>% group_by(phq9_9) %>%
  summarise(frequency = n())
```

```
## # A tibble: 2 x 2
##   phq9_9 frequency
##    <int>     <int>
## 1      0       268
## 2      1         4
```

```
data_filter %>% ggplot(aes(x = phq9_9, y = phq_9_score)) + geom_count()
```




```
#test whether phq9_9 is related to ph9_9_score using t-test

t.test(data_filter[data_filter$phq9_9 == 1,]$phq_9_score,data_filter[data_filter$phq9_9 == 0,]$phq_9_sc
```

```
##
##  Welch Two Sample t-test
##
## data:  data_filter[data_filter$phq9_9 == 1, ]$phq_9_score and data_filter[data_filter$phq9_9 == 0, ]$
## t = 2.6661, df = 3.0271, p-value = 0.07522
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.053732 12.285076
## sample estimates:
## mean of x mean of y
##  7.500000  1.884328
```

```
## Explore each question of PHQ9

for (i in 1:9) {
  print(paste0("PHQ9_",i,": ",
               attributes((data_filter[,32:40])[,i,drop = TRUE])$label))
}
```

```
## [1] "PHQ9_1: Little interest or pleasure in doing things"
## [1] "PHQ9_2: Feeling down, depressed, or hopeless"
```

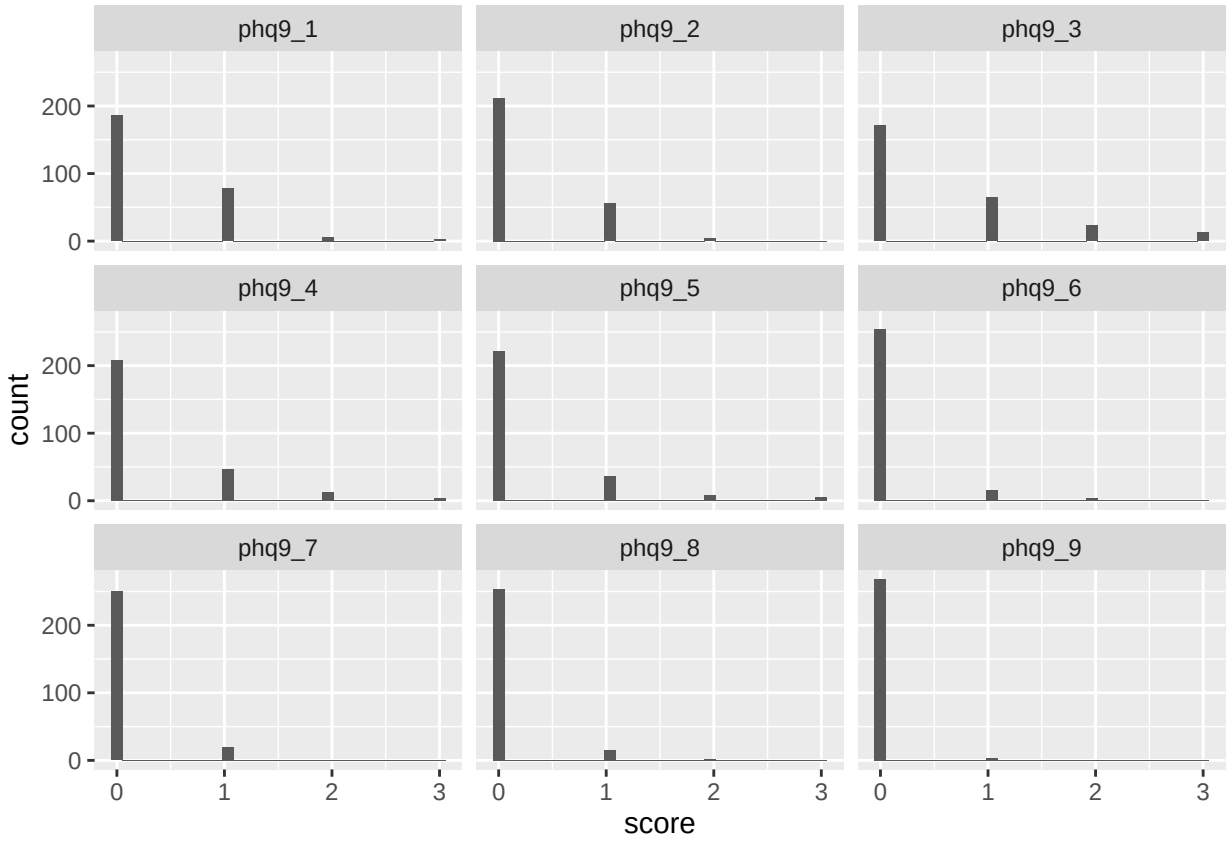
```
## [1] "PHQ9_3: Trouble falling or staying asleep, or sleeping too much"
## [1] "PHQ9_4: Feeling tired or having little energy"
## [1] "PHQ9_5: Poor appetite or overeating"
## [1] "PHQ9_6: Feeling bad about yourself - or that you are a failure or have let yourself or your fami
## [1] "PHQ9_7: Trouble concentrating on things, such as reading the newspaper or watching television"
## [1] "PHQ9_8: Moving or speaking so slowly that other people could have noticed? Or so fidgety or rest
## [1] "PHQ9_9: Thoughts that you would be better off dead, or thoughts of hurting yourself in some way
```

```
#convert wide to long to be used in ggplot
data_long <- gather(data_filter, phq, score, phq9_1:phq9_9, factor_key=TRUE)
```

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

```
data_long %>% ggplot(aes(x=score)) + geom_histogram() + facet_wrap(~phq)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



#Check before analyisis

```
#Check for Multicollinearity
```

```r
column_to_excludes <- c("WeightofPatient","HeightofPatient",
                        "phq9_1","phq9_2","phq9_3","phq9_4","phq9_5","phq9_6","phq9_7","phq9_8","phq9_9

mix_assoc_result <- mixed_assoc(data_filter[,!colnames(data_filter) %in% column_to_excludes])
```

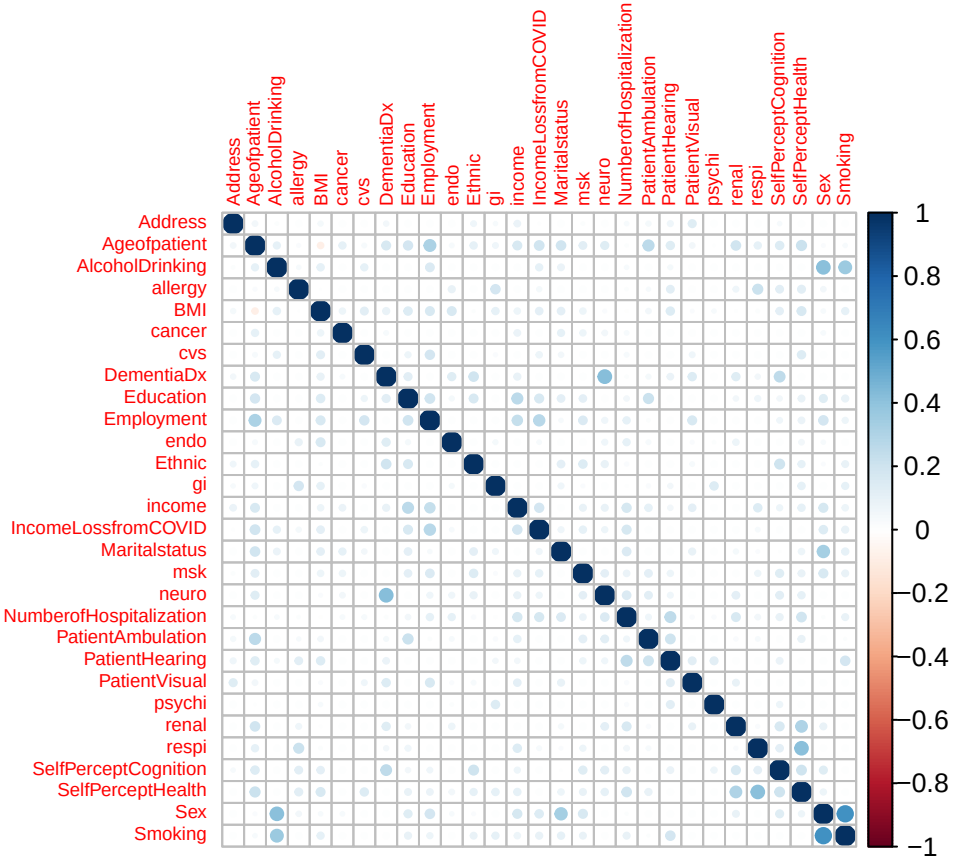
```r
mix_assoc_result %>%
  dplyr::select(-complete_obs_pairs, -complete_obs_ratio) %>%
  filter(x != y) %>%
  filter(assoc != 0) %>%
  arrange(desc(abs(assoc))) %>%
  filter(row_number() %% 2 == 1) %>%
  group_by(type) %>%
  slice_max(order_by = abs(assoc), n = 10) %>%
  ungroup()
```

```
## # A tibble: 23 x 4
##    x                      y                      assoc type
##    <chr>                  <chr>                  <dbl> <chr>
##  1 Employment             Ageofpatient           0.312 anova
##  2 PatientAmbulation      Ageofpatient           0.269 anova
##  3 NumberofHospitalization PatientHearing        0.259 anova
##  4 SelfPerceptHealth      Ageofpatient           0.219 anova
##  5 IncomeLossfromCOVID    Ageofpatient           0.199 anova
##  6 SelfPerceptHealth      NumberofHospitalization 0.198 anova
##  7 renal                  Ageofpatient           0.190 anova
##  8 Maritalstatus          Ageofpatient           0.190 anova
##  9 Education              Ageofpatient           0.180 anova
## 10 NumberofHospitalization income                0.177 anova
## # i 13 more rows
```

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
mix_assoc_result %>%
    dplyr::select(x,y,assoc) %>%
    spread(y, assoc) %>%
    column_to_rownames("x") %>%
    as.matrix %>%
    corrplot(tl.cex = 0.6)
```




```
#explore vif & gvif

column_to_excludes <- c("WeightofPatient","HeightofPatient",
                        "phq9_1","phq9_2","phq9_3","phq9_4","phq9_5","phq9_6","phq9_7","phq9_8","phq9_9


lm_model <- lm(phq_9_score ~ ., data = data_filter[,!colnames(data_filter) %in% column_to_excludes])

vif(lm_model)
```

```
##                              GVIF Df GVIF^(1/(2*Df))
## Ageofpatient             1.617595  1        1.271847
## Sex                      2.506254  1        1.583115
## BMI                      1.334822  1        1.155345
## Ethnic                   1.361781  1        1.166954
## Maritalstatus            1.990436  3        1.121566
## Address                  1.197630  1        1.094363
## Education                2.761617  3        1.184479
## Employment               3.056393  3        1.204670
## income                   3.488635  4        1.169047
## IncomeLossfromCOVID      2.516966  3        1.166307
## PatientAmbulation        1.376356  1        1.173182
## PatientHearing           1.828684  2        1.162879
## PatientVisual            1.315209  1        1.146826
## Smoking                  2.806533  2        1.294323
## AlcoholDrinking          1.861722  2        1.168096
```

```
## DementiaDx              2.066491  2        1.198970
## SelfPerceptCognition    1.947143  2        1.181271
## NumberofHospitalization 1.406476  1        1.185950
## SelfPerceptHealth       2.992895  4        1.146863
## neuro                   1.495258  1        1.222807
## cvs                     1.244200  1        1.115437
## respi                   1.446223  1        1.202590
## gi                      1.245078  1        1.115831
## renal                   1.355832  1        1.164402
## endo                    1.246383  1        1.116415
## msk                     1.344023  1        1.159320
## cancer                  1.189774  1        1.090768
## allergy                 1.288468  1        1.135107
## psychi                  1.168477  1        1.080961
```

```r
## Try the automatic selection by collinear
library(collinear)
column_to_excludes_initial <- c("WeightofPatient","HeightofPatient",
                        "phq9_1","phq9_2","phq9_3","phq9_4","phq9_5","phq9_6","phq9_7","phq9_8","phq9_9

response_choices <- c("phq_9_cat","phq_9_score")
selected_response <- response_choices[2]

column_to_excludes <- setdiff(column_to_excludes_initial,selected_response)

predictors <- setdiff(setdiff(colnames(data_filter),column_to_excludes),selected_response)
```

```r
#for linear
selected_predictors_no_preference <- collinear(
  df = data_filter[,!colnames(data_filter) %in% column_to_excludes],
  response = selected_response,
  predictors = predictors,
  preference_order = NULL,
  max_cor = 0.5,
  max_vif = 2.5,
  encoding_method = "mean"
)


preference_rsquared <- preference_order(
  df = data_filter[,!colnames(data_filter) %in% column_to_excludes],
  response = selected_response,
  predictors = predictors,
  f = f_rsquared,
  workers = 4 #requires package future and future.apply for more workers
)

selected_predictors_with_preference <- collinear(
  df = data_filter[,!colnames(data_filter) %in% column_to_excludes],
  response = selected_response,
  predictors = predictors,
  preference_order = preference_rsquared,
  max_cor = 0.5,
  max_vif = 2.5,
```

```
  encoding_method = "mean"
)
```

```
selected_predictors_no_preference
```

```
##  [1] "AlcoholDrinking"         "cancer"
##  [3] "psychi"                  "cvs"
##  [5] "Address"                 "PatientVisual"
##  [7] "Maritalstatus"           "Smoking"
##  [9] "PatientHearing"          "NumberofHospitalization"
## [11] "endo"                    "gi"
## [13] "BMI"                     "Ethnic"
## [15] "msk"                     "PatientAmbulation"
## [17] "allergy"                 "renal"
## [19] "IncomeLossfromCOVID"     "Employment"
## [21] "SelfPerceptCognition"    "respi"
## [23] "Ageofpatient"            "neuro"
## [25] "Education"               "income"
## [27] "DementiaDx"              "SelfPerceptHealth"
```

```
selected_predictors_with_preference
```

```
##  [1] "SelfPerceptHealth"       "SelfPerceptCognition"
##  [3] "NumberofHospitalization" "gi"
##  [5] "renal"                   "msk"
##  [7] "neuro"                   "respi"
##  [9] "psychi"                  "DementiaDx"
## [11] "allergy"                 "PatientHearing"
## [13] "income"                  "Ageofpatient"
## [15] "IncomeLossfromCOVID"     "Education"
## [17] "Maritalstatus"           "Employment"
## [19] "endo"                    "Address"
## [21] "Ethnic"                  "PatientAmbulation"
## [23] "cancer"                  "AlcoholDrinking"
## [25] "PatientVisual"           "BMI"
## [27] "Sex"                     "cvs"
```

```
#for logistic

selected_predictors_no_response <- cor_select(
  df = data_filter[,!colnames(data_filter) %in% column_to_excludes],
  predictors = predictors,
  preference_order = preference_rsquared,
  max_cor = 0.5
)

selected_predictors_no_response
```

```
##  [1] "SelfPerceptHealth"       "SelfPerceptCognition"
##  [3] "NumberofHospitalization" "gi"
##  [5] "renal"                   "msk"
##  [7] "neuro"                   "respi"
```

```
##  [9] "psychi"                  "DementiaDx"
## [11] "allergy"                 "PatientHearing"
## [13] "income"                  "Ageofpatient"
## [15] "IncomeLossfromCOVID"     "Education"
## [17] "Maritalstatus"           "Employment"
## [19] "endo"                    "Address"
## [21] "Ethnic"                  "PatientAmbulation"
## [23] "cancer"                  "AlcoholDrinking"
## [25] "PatientVisual"           "BMI"
## [27] "Sex"                     "cvs"
```

```
# Exclude specified columns
column_to_excludes <- c("WeightofPatient", "HeightofPatient",
                        "phq9_1", "phq9_2", "phq9_3", "phq9_4", "phq9_5", "phq9_6", "phq9_7", "phq9_8",
full_var <- setdiff(colnames(data_filter), column_to_excludes)

# Create the initial data frame
df_collinear_remove <- data.frame(original = full_var,
                        collinear_linear_with_preference = full_var,
                        collinear_linear_no_preference = full_var,
                        collinear_no_response = full_var)

# Helper function to replace variables with NA
replace_with_na <- function(df, col_name, exclude_vars) {
  vars_to_replace <- setdiff(full_var, exclude_vars)
  df[[col_name]] <- sapply(df[[col_name]], function(x) if (x %in% vars_to_replace) NA_character_ else x)
  return(df)
}

# Apply the helper function for each scenario
df_collinear_remove <- replace_with_na(df_collinear_remove, "collinear_linear_with_preference", selected
df_collinear_remove <- replace_with_na(df_collinear_remove, "collinear_linear_no_preference", selected_p
df_collinear_remove <- replace_with_na(df_collinear_remove, "collinear_no_response", selected_predictors
```

```
df_collinear_remove %>%
  filter(if_any(everything(), is.na))
```

```
##   original collinear_linear_with_preference collinear_linear_no_preference
## 1      Sex                              Sex                            <NA>
## 2  Smoking                             <NA>                         Smoking
##   collinear_no_response
## 1                   Sex
## 2                  <NA>
```

##Multiple linear regression

```
# Multiple linear regression
column_to_excludes <- c("WeightofPatient","HeightofPatient",
                        "phq9_1","phq9_2","phq9_3","phq9_4","phq9_5","phq9_6","phq9_7","phq9_8","phq9_9


lm_model <- lm(phq_9_score ~ ., data = data_filter[,!colnames(data_filter) %in% column_to_excludes])
```

```r
# Print the summary of the model
summary(lm_model)
```

```
##
## Call:
## lm(formula = phq_9_score ~ ., data = data_filter[, !colnames(data_filter) %in%
##     column_to_excludes])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7358 -1.1611 -0.2588  0.9401  7.5631
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                  0.24265    3.27804   0.074
## Ageofpatient                                 0.01848    0.02096   0.882
## Sexfemale                                    0.29828    0.43148   0.691
## BMI                                          0.02409    0.03821   0.630
## EthnicChinese                               -0.16502    0.80418  -0.205
## Maritalstatusmarried                         0.22474    0.37022   0.607
## Maritalstatusdivorced                        1.09131    0.76230   1.432
## Maritalstatuswidow                           0.16666    0.47880   0.348
## Addressothers                               -0.37784    0.28668  -1.318
## Educationelementary                          0.69350    1.34605   0.515
## Educationhigh school                         0.85522    1.34456   0.636
## Educationcollege degree                      0.67651    1.34328   0.504
## Employmentpart-time job                      0.40611    0.58604   0.693
## Employmentfull-time job                     -0.23396    0.57241  -0.409
## Employmentretired                            0.09613    0.33066   0.291
## income10,001 - 20,000                       -0.01459    0.49173  -0.030
## income20,001 - 30,000                       -0.49120    0.53200  -0.923
## income30,001 or more                        -0.28644    0.50305  -0.569
## incomeunknown                               -0.59329    0.46616  -1.273
## IncomeLossfromCOVIDLess than 50% loss        0.44294    0.56331   0.786
## IncomeLossfromCOVIDOver 50% loss             0.69326    0.59942   1.157
## IncomeLossfromCOVIDNo income                 0.31642    0.67508   0.469
## PatientAmbulationGait aid                   -0.46856    0.70418  -0.665
## PatientHearingHearing aid                   -0.27067    1.74706  -0.155
## PatientHearingHearing impairment             0.98130    0.60544   1.621
## PatientVisualGlasses                        -0.29620    0.29318  -1.010
## SmokingCurrent smoking                      -0.48725    1.05172  -0.463
## SmokingPast smoking                          0.16592    0.53231   0.312
## AlcoholDrinkingSocial drinking               0.40957    0.55713   0.735
## AlcoholDrinkingRegular drinking              1.36443    1.31827   1.035
## DementiaDxYes                                0.52525    0.80805   0.650
## DementiaDxNot sure                           1.11531    0.92357   1.208
## SelfPerceptCognitionMinor cognitive problem  0.25092    0.28735   0.873
## SelfPerceptCognitionMajor cognitive problem  2.33222    1.75707   1.327
## NumberofHospitalization                      0.70090    0.30270   2.316
## SelfPerceptHealthBad                         1.54398    2.46422   0.627
## SelfPerceptHealthAverage                    -1.73296    2.19174  -0.791
## SelfPerceptHealthGood                       -2.46493    2.18391  -1.129
## SelfPerceptHealthBest                       -2.89334    2.23010  -1.297
```

```
## neuroNeurological disease                    0.44348    0.45575   0.973
## cvsCardiovascular disease                   -0.23949    0.34712  -0.690
## respiRespiratory disease                     0.28596    0.55535   0.515
## giGastrointestinal disease                   0.96134    0.34015   2.826
## renalRenal disease                           0.51810    0.52641   0.984
## endoEndocrine disease                        0.14106    0.32993   0.428
## mskMSK disease                               0.82064    0.29637   2.769
## cancerCancer                                 0.39821    0.48324   0.824
## allergyAllergy                               0.51988    0.31309   1.660
## psychiPsych disease                          1.66403    0.58371   2.851
##                                             Pr(>|t|)
## (Intercept)                                  0.94106
## Ageofpatient                                 0.37890
## Sexfemale                                    0.49010
## BMI                                          0.52908
## EthnicChinese                                0.83760
## Maritalstatusmarried                         0.54443
## Maritalstatusdivorced                        0.15366
## Maritalstatuswidow                           0.72811
## Addressothers                                0.18886
## Educationelementary                          0.60692
## Educationhigh school                         0.52539
## Educationcollege degree                      0.61502
## Employmentpart-time job                      0.48905
## Employmentfull-time job                      0.68313
## Employmentretired                            0.77154
## income10,001 - 20,000                        0.97635
## income20,001 - 30,000                        0.35684
## income30,001 or more                         0.56966
## incomeunknown                                0.20445
## IncomeLossfromCOVIDLess than 50% loss        0.43251
## IncomeLossfromCOVIDOver 50% loss             0.24870
## IncomeLossfromCOVIDNo income                 0.63974
## PatientAmbulationGait aid                    0.50649
## PatientHearingHearing aid                    0.87702
## PatientHearingHearing impairment             0.10647
## PatientVisualGlasses                         0.31345
## SmokingCurrent smoking                       0.64361
## SmokingPast smoking                          0.75556
## AlcoholDrinkingSocial drinking               0.46303
## AlcoholDrinkingRegular drinking              0.30178
## DementiaDxYes                                0.51635
## DementiaDxNot sure                           0.22848
## SelfPerceptCognitionMinor cognitive problem  0.38347
## SelfPerceptCognitionMajor cognitive problem  0.18575
## NumberofHospitalization                      0.02149 *
## SelfPerceptHealthBad                         0.53159
## SelfPerceptHealthAverage                     0.42997
## SelfPerceptHealthGood                        0.26025
## SelfPerceptHealthBest                        0.19583
## neuroNeurological disease                    0.33157
## cvsCardiovascular disease                    0.49096
## respiRespiratory disease                     0.60712
## giGastrointestinal disease                   0.00514 **
```

```
## renalRenal disease                              0.32607
## endoEndocrine disease                           0.66940
## mskMSK disease                                  0.00610 **
## cancerCancer                                    0.41080
## allergyAllergy                                  0.09823 .
## psychiPsych disease                             0.00477 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.033 on 223 degrees of freedom
## Multiple R-squared:  0.425,  Adjusted R-squared:  0.3012
## F-statistic: 3.434 on 48 and 223 DF,  p-value: 2.991e-10
```

```
lm_tbl <- lm_model %>% tbl_regression()
lm_tbl
```

Characteristic	**Beta**	**95% CI**	**p-value**
Age	0.02	-0.02, 0.06	0.4
Sex			
male	—	—	
female	0.30	-0.55, 1.1	0.5
BMI	0.02	-0.05, 0.10	0.5
Ethnic			
Thai	—	—	
Chinese	-0.17	-1.7, 1.4	0.8
Marital status			
single	—	—	
married	0.22	-0.50, 0.95	0.5
divorced	1.1	-0.41, 2.6	0.2
widow	0.17	-0.78, 1.1	0.7
Address			
Bangkok	—	—	
others	-0.38	-0.94, 0.19	0.2
Education			
not educate	—	—	
elementary	0.69	-2.0, 3.3	0.6
high school	0.86	-1.8, 3.5	0.5
college degree	0.68	-2.0, 3.3	0.6
Employment			
unemployed	—	—	
part-time job	0.41	-0.75, 1.6	0.5
full-time job	-0.23	-1.4, 0.89	0.7
retired	0.10	-0.56, 0.75	0.8
income			
10,000 or less	—	—	
10,001 - 20,000	-0.01	-0.98, 0.95	>0.9
20,001 - 30,000	-0.49	-1.5, 0.56	0.4
30,001 or more	-0.29	-1.3, 0.70	0.6
unknown	-0.59	-1.5, 0.33	0.2
Income Loss from COVID-19			
Same	—	—	
Less than 50% loss	0.44	-0.67, 1.6	0.4

Characteristic	Beta	95% CI	p-value
Over 50% loss	0.69	-0.49, 1.9	0.2
No income	0.32	-1.0, 1.6	0.6
PatientAmbulation			
Normal	—	—	
Gait aid	-0.47	-1.9, 0.92	0.5
Hearing			
Normal	—	—	
Hearing aid	-0.27	-3.7, 3.2	0.9
Hearing impairment	0.98	-0.21, 2.2	0.11
PatientVisual			
Normal	—	—	
Glasses	-0.30	-0.87, 0.28	0.3
Smoking			
Never smoking	—	—	
Current smoking	-0.49	-2.6, 1.6	0.6
Past smoking	0.17	-0.88, 1.2	0.8
Alcohol Drinking			
Never drinking	—	—	
Social drinking	0.41	-0.69, 1.5	0.5
Regular drinking	1.4	-1.2, 4.0	0.3
Dementia diagnosis			
No	—	—	
Yes	0.53	-1.1, 2.1	0.5
Not sure	1.1	-0.70, 2.9	0.2
Self Percept Cognition			
Normal	—	—	
Minor cognitive problem	0.25	-0.32, 0.82	0.4
Major cognitive problem	2.3	-1.1, 5.8	0.2
Number of Hospitalization	0.70	0.10, 1.3	0.021
Self Percept Health			
Worst	—	—	
Bad	1.5	-3.3, 6.4	0.5
Average	-1.7	-6.1, 2.6	0.4
Good	-2.5	-6.8, 1.8	0.3
Best	-2.9	-7.3, 1.5	0.2
neuro			
None	—	—	
Neurological disease	0.44	-0.45, 1.3	0.3
cvs			
None	—	—	
Cardiovascular disease	-0.24	-0.92, 0.44	0.5
respi			
None	—	—	
Respiratory disease	0.29	-0.81, 1.4	0.6
gi			
None	—	—	
Gastrointestinal disease	0.96	0.29, 1.6	0.005
renal			
None	—	—	
Renal disease	0.52	-0.52, 1.6	0.3
endo			
None	—	—	

Characteristic	Beta	95% CI	p-value
Endocrine disease	0.14	-0.51, 0.79	0.7
msk			
None	—	—	
MSK disease	0.82	0.24, 1.4	0.006
cancer			
None	—	—	
Cancer	0.40	-0.55, 1.4	0.4
allergy			
None	—	—	
Allergy	0.52	-0.10, 1.1	0.10
psychi			
None	—	—	
Psych disease	1.7	0.51, 2.8	0.005

Ordered logistic regression

```
# Ordered logistic regression
column_to_excludes <- c("WeightofPatient","HeightofPatient",
                        "phq9_1","phq9_2","phq9_3","phq9_4","phq9_5","phq9_6","phq9_7","phq9_8","phq9_9



ordered_logistic_model <- polr(phq_9_cat ~ ., data = data_filter[,!colnames(data_filter) %in% column_to
```

```
## Warning in polr(phq_9_cat ~ ., data = data_filter[, !colnames(data_filter) %in%
## : design appears to be rank-deficient, so dropping some coefs
```

```
# Print the summary of the model
summary(ordered_logistic_model)
```

```
## Call:
## polr(formula = phq_9_cat ~ ., data = data_filter[, !colnames(data_filter) %in%
##     column_to_excludes], Hess = TRUE)
##
## Coefficients:
##                                          Value Std. Error    t value
## Ageofpatient                          0.072717  4.146e-02  1.754e+00
## Sexfemale                            -0.443665  8.675e-01 -5.115e-01
## BMI                                   0.061498  7.057e-02  8.714e-01
## EthnicChinese                        -1.178456  1.331e+00 -8.852e-01
## Maritalstatusmarried                 -0.288887  6.944e-01 -4.160e-01
## Maritalstatusdivorced                 1.336868  1.224e+00  1.092e+00
## Maritalstatuswidow                    0.001621  8.000e-01  2.026e-03
## Addressothers                        -1.219883  6.249e-01 -1.952e+00
## Educationelementary                  -0.336985  1.730e+00 -1.947e-01
## Educationhigh school                 -0.459330  1.614e+00 -2.847e-01
## Educationcollege degree              -0.545220  1.619e+00 -3.367e-01
## Employmentpart-time job               0.009879  9.408e-01  1.050e-02
```

```
## Employmentfull-time job                          0.014816  1.108e+00  1.337e-02
## Employmentretired                               -0.237262  5.723e-01 -4.146e-01
## income10,001 - 20,000                           -0.026576  7.994e-01 -3.325e-02
## income20,001 - 30,000                           -0.719888  1.034e+00 -6.965e-01
## income30,001 or more                            -0.755314  9.089e-01 -8.310e-01
## incomeunknown                                   -0.071800  7.160e-01 -1.003e-01
## IncomeLossfromCOVIDLess than 50% loss            1.100814  8.331e-01  1.321e+00
## IncomeLossfromCOVIDOver 50% loss                 0.518023  8.998e-01  5.757e-01
## IncomeLossfromCOVIDNo income                     0.979288  1.089e+00  8.992e-01
## PatientAmbulationGait aid                       -1.263488  1.009e+00 -1.252e+00
## PatientHearingHearing aid                      -14.590159  7.729e-07 -1.888e+07
## PatientHearingHearing impairment                 1.414797  9.027e-01  1.567e+00
## PatientVisualGlasses                            -0.765627  5.692e-01 -1.345e+00
## SmokingCurrent smoking                          -0.404857  1.728e+00 -2.343e-01
## SmokingPast smoking                              0.325965  9.645e-01  3.380e-01
## AlcoholDrinkingSocial drinking                   0.534819  9.641e-01  5.548e-01
## AlcoholDrinkingRegular drinking                  1.642117  1.594e+00  1.030e+00
## DementiaDxYes                                    0.248869  1.065e+00  2.338e-01
## DementiaDxNot sure                               2.146607  1.141e+00  1.881e+00
## SelfPerceptCognitionMinor cognitive problem      0.665308  5.877e-01  1.132e+00
## SelfPerceptCognitionMajor cognitive problem      1.736210  2.114e+00  8.214e-01
## NumberofHospitalization                          0.790653  4.199e-01  1.883e+00
## SelfPerceptHealthBad                            -3.172385  2.844e+00 -1.115e+00
## SelfPerceptHealthAverage                        -4.315184  2.374e+00 -1.818e+00
## SelfPerceptHealthGood                           -5.195023  2.356e+00 -2.205e+00
## SelfPerceptHealthBest                           -5.506906  2.644e+00 -2.083e+00
## neuroNeurological disease                        0.360566  6.552e-01  5.503e-01
## cvsCardiovascular disease                        0.279703  6.742e-01  4.149e-01
## respiRespiratory disease                         0.847923  8.878e-01  9.550e-01
## giGastrointestinal disease                       0.917673  5.483e-01  1.674e+00
## renalRenal disease                               0.442545  7.669e-01  5.770e-01
## endoEndocrine disease                           -0.402081  5.900e-01 -6.814e-01
## mskMSK disease                                   1.242114  5.383e-01  2.307e+00
## cancerCancer                                    -0.080826  7.740e-01 -1.044e-01
## allergyAllergy                                  -0.183143  5.920e-01 -3.094e-01
## psychiPsych disease                              2.044062  7.632e-01  2.678e+00
##
## Intercepts:
##                                        Value         Std. Error    t value
## normal|mild depression                  4.525300e+00  4.680200e+00  9.669000e-01
## mild depression|moderate depression     7.332900e+00  4.706400e+00  1.558100e+00
##
## Residual Deviance: 177.9185
## AIC: 277.9185
```

```
ordered_logistic_tbl <- ordered_logistic_model %>% tbl_regression(exponentiate = TRUE)
ordered_logistic_tbl
```

Characteristic	OR	95% CI	p-value
Age	1.08	0.99, 1.17	0.081
Sex			
male	—	—	

Characteristic	OR	95% CI	p-value
female	0.64	0.12, 3.55	0.6
BMI	1.06	0.93, 1.22	0.4
Ethnic			
Thai	—	—	
Chinese	0.31	0.02, 4.24	0.4
Marital status			
single	—	—	
married	0.75	0.19, 2.94	0.7
divorced	3.81	0.34, 42.5	0.3
widow	1.00	0.21, 4.85	>0.9
Address			
Bangkok	—	—	
others	0.30	0.09, 1.01	0.052
Education			
not educate	—	—	
elementary	0.71	0.02, 21.6	0.8
high school	0.63	0.03, 15.2	0.8
college degree	0.58	0.02, 14.1	0.7
Employment			
unemployed	—	—	
part-time job	1.01	0.16, 6.45	>0.9
full-time job	1.01	0.11, 9.01	>0.9
retired	0.79	0.26, 2.44	0.7
income			
10,000 or less	—	—	
10,001 - 20,000	0.97	0.20, 4.71	>0.9
20,001 - 30,000	0.49	0.06, 3.73	0.5
30,001 or more	0.47	0.08, 2.82	0.4
unknown	0.93	0.23, 3.82	>0.9
Income Loss from COVID-19			
Same	—	—	
Less than 50% loss	3.01	0.58, 15.5	0.2
Over 50% loss	1.68	0.29, 9.89	0.6
No income	2.66	0.31, 22.8	0.4
Ambulation			
Normal	—	—	
Gait aid	0.28	0.04, 2.07	0.2
Hearing			
Normal	—	—	
Hearing aid	0.00	0.00, 0.00	<0.001
Hearing impairment	4.12	0.69, 24.4	0.12
Visual			
Normal	—	—	
Glasses	0.47	0.15, 1.43	0.2
Smoking			
Never smoking	—	—	
Current smoking	0.67	0.02, 20.1	0.8
Past smoking	1.39	0.21, 9.27	0.7
Alcohol Drinking			
Never drinking	—	—	
Social drinking	1.71	0.26, 11.4	0.6
Regular drinking	5.17	0.22, 119	0.3

Characteristic	OR	95% CI	p-value
Dementia diagnosis			
No	—	—	
Yes	1.28	0.16, 10.5	0.8
Not sure	8.56	0.90, 81.1	0.061
Self Percept Cognition			
Normal	—	—	
Minor cognitive problem	1.95	0.61, 6.19	0.3
Major cognitive problem	5.68	0.09, 366	0.4
Number of Hospitalization	2.20	0.96, 5.04	0.061
Self Percept Health			
Worst	—	—	
Bad	0.04	0.00, 11.4	0.3
Average	0.01	0.00, 1.44	0.070
Good	0.01	0.00, 0.58	0.028
Best	0.00	0.00, 0.74	0.038
neuro			
None	—	—	
Neurological disease	1.43	0.39, 5.22	0.6
cvs			
None	—	—	
Cardiovascular disease	1.32	0.35, 4.99	0.7
respi			
None	—	—	
Respiratory disease	2.33	0.41, 13.4	0.3
gi			
None	—	—	
Gastrointestinal disease	2.50	0.85, 7.38	0.10
renal			
None	—	—	
Renal disease	1.56	0.34, 7.06	0.6
endo			
None	—	—	
Endocrine disease	0.67	0.21, 2.14	0.5
msk			
None	—	—	
MSK disease	3.46	1.20, 10.0	0.022
cancer			
None	—	—	
Cancer	0.92	0.20, 4.24	>0.9
allergy			
None	—	—	
Allergy	0.83	0.26, 2.67	0.8
psychi			
None	—	—	
Psych disease	7.72	1.72, 34.7	0.008

```
Anova(ordered_logistic_model)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: phq_9_cat
```

```
##                          LR Chisq Df Pr(>Chisq)
## Ageofpatient               3.2004  1   0.073621 .
## Sex                        0.2572  1   0.612027
## BMI                        0.7566  1   0.384388
## Ethnic                     0.8341  1   0.361085
## Maritalstatus              2.0354  3   0.565089
## Address                    4.2501  1   0.039247 *
## Education                  0.1671  4   0.996696
## Employment                 0.2004  3   0.977528
## income                     1.4045  4   0.843419
## IncomeLossfromCOVID        2.0879  3   0.554363
## PatientAmbulation          1.6674  2   0.434449
## PatientHearing             2.8932  2   0.235365
## PatientVisual              1.9091  2   0.384993
## Smoking                    0.2596  2   0.878251
## AlcoholDrinking            1.1478  2   0.563321
## DementiaDx                 3.4488  2   0.178281
## SelfPerceptCognition       1.8521  2   0.396105
## NumberofHospitalization    3.3128  1   0.068744 .
## SelfPerceptHealth          9.2868  4   0.054318 .
## neuro                      0.2985  1   0.584840
## cvs                        0.1752  1   0.675539
## respi                      0.8723  1   0.350311
## gi                         2.7329  1   0.098300 .
## renal                      0.3232  1   0.569686
## endo                       0.4796  1   0.488602
## msk                        5.6402  1   0.017553 *
## cancer                     0.0110  1   0.916400
## allergy                    0.0968  1   0.755694
## psychi                     6.7703  1   0.009269 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Binary logistic regression

```r
#Binary logistic regression only for the phq9-9 question
column_to_excludes <- c("WeightofPatient","HeightofPatient",
                        "phq9_1","phq9_2","phq9_3","phq9_4","phq9_5","phq9_6","phq9_7","phq9_8","phq_9_s

column_to_excludes <- c("WeightofPatient","HeightofPatient","phq_9_score","phq_9_cat")

binary_logistic_model <- glm(phq9_9 ~ ., family = binomial(), data = data_filter[,!colnames(data_filter)

summary(binary_logistic_model)
```

```
##
## Call:
## glm(formula = phq9_9 ~ ., family = binomial(), data = data_filter[,
##     !colnames(data_filter) %in% column_to_excludes])
##
## Coefficients:
##                                          Estimate Std. Error z value
## (Intercept)                            -6.035e+01  7.555e+05       0
```

```
## Ageofpatient                                  7.525e-02  5.154e+03       0
## Sexfemale                                    -4.997e-01  1.356e+05       0
## BMI                                           3.560e-01  1.177e+04       0
## EthnicChinese                                 1.738e+01  2.042e+05       0
## Maritalstatusmarried                         -9.847e+00  8.048e+04       0
## Maritalstatusdivorced                         3.766e+00  1.402e+05       0
## Maritalstatuswidow                           -1.225e+01  1.221e+05       0
## Addressothers                                 3.728e+00  8.157e+04       0
## Educationelementary                           1.518e+00  3.711e+05       0
## Educationhigh school                         -7.278e+00  4.199e+05       0
## Educationcollege degree                      -1.096e+00  4.177e+05       0
## Employmentpart-time job                      -1.363e+00  2.197e+05       0
## Employmentfull-time job                       8.157e+00  1.288e+05       0
## Employmentretired                             2.942e+00  1.209e+05       0
## income10,001 - 20,000                        -1.548e-01  1.502e+05       0
## income20,001 - 30,000                         8.566e+00  1.551e+05       0
## income30,001 or more                          6.771e+00  1.675e+05       0
## incomeunknown                                -4.941e+00  1.608e+05       0
## IncomeLossfromCOVIDLess than 50% loss         1.713e+01  8.553e+04       0
## IncomeLossfromCOVIDOver 50% loss              6.841e-01  2.039e+05       0
## IncomeLossfromCOVIDNo income                 -9.235e+00  2.407e+05       0
## PatientAmbulationGait aid                     9.908e+00  2.117e+05       0
## PatientHearingHearing aid                     6.833e+00  4.145e+05       0
## PatientHearingHearing impairment             -8.416e+00  2.820e+05       0
## PatientVisualGlasses                         -1.665e+00  1.013e+05       0
## SmokingCurrent smoking                        1.680e+01  1.553e+05       0
## SmokingPast smoking                          -9.472e-01  1.639e+05       0
## AlcoholDrinkingSocial drinking                7.591e+00  1.158e+05       0
## AlcoholDrinkingRegular drinking               3.673e+01  1.998e+05       0
## DementiaDxYes                                -1.232e+01  2.520e+05       0
## DementiaDxNot sure                           -2.577e+01  3.136e+05       0
## SelfPerceptCognitionMinor cognitive problem   5.849e+00  9.107e+04       0
## SelfPerceptCognitionMajor cognitive problem   4.150e+01  3.738e+05       0
## NumberofHospitalization                       4.106e+00  8.478e+04       0
## SelfPerceptHealthBad                         -1.519e+00  6.602e+05       0
## SelfPerceptHealthAverage                      1.980e+00  5.151e+05       0
## SelfPerceptHealthGood                         5.140e+00  5.144e+05       0
## SelfPerceptHealthBest                         3.348e+00  4.937e+05       0
## neuroNeurological disease                     6.722e+00  8.766e+04       0
## cvsCardiovascular disease                     5.029e+00  1.116e+05       0
## respiRespiratory disease                      5.525e+00  1.610e+05       0
## giGastrointestinal disease                    1.736e+00  9.410e+04       0
## renalRenal disease                           -2.198e+00  2.022e+05       0
## endoEndocrine disease                         2.464e+00  1.072e+05       0
## mskMSK disease                               -5.005e+00  7.537e+04       0
## cancerCancer                                  1.198e+01  8.655e+04       0
## allergyAllergy                               -1.466e+00  7.280e+04       0
## psychiPsych disease                          -7.514e+00  1.772e+05       0
## phq9_1                                       -4.660e-01  9.683e+04       0
## phq9_2                                        5.605e+00  8.495e+04       0
## phq9_3                                       -3.003e+00  6.235e+04       0
## phq9_4                                        5.273e+00  6.380e+04       0
## phq9_5                                       -9.714e-01  5.778e+04       0
## phq9_6                                       -1.110e+00  1.617e+05       0
```

```
## phq9_7                                    -8.972e+00  2.451e+05       0
## phq9_8                                     7.688e+00  1.471e+05       0
##                                           Pr(>|z|)
## (Intercept)                                      1
## Ageofpatient                                     1
## Sexfemale                                        1
## BMI                                              1
## EthnicChinese                                    1
## Maritalstatusmarried                             1
## Maritalstatusdivorced                            1
## Maritalstatuswidow                               1
## Addressothers                                    1
## Educationelementary                              1
## Educationhigh school                             1
## Educationcollege degree                          1
## Employmentpart-time job                          1
## Employmentfull-time job                          1
## Employmentretired                                1
## income10,001 - 20,000                            1
## income20,001 - 30,000                            1
## income30,001 or more                             1
## incomeunknown                                    1
## IncomeLossfromCOVIDLess than 50% loss            1
## IncomeLossfromCOVIDOver 50% loss                 1
## IncomeLossfromCOVIDNo income                     1
## PatientAmbulationGait aid                        1
## PatientHearingHearing aid                        1
## PatientHearingHearing impairment                 1
## PatientVisualGlasses                             1
## SmokingCurrent smoking                           1
## SmokingPast smoking                              1
## AlcoholDrinkingSocial drinking                   1
## AlcoholDrinkingRegular drinking                  1
## DementiaDxYes                                    1
## DementiaDxNot sure                               1
## SelfPerceptCognitionMinor cognitive problem      1
## SelfPerceptCognitionMajor cognitive problem      1
## NumberofHospitalization                          1
## SelfPerceptHealthBad                             1
## SelfPerceptHealthAverage                         1
## SelfPerceptHealthGood                            1
## SelfPerceptHealthBest                            1
## neuroNeurological disease                        1
## cvsCardiovascular disease                        1
## respiRespiratory disease                         1
## giGastrointestinal disease                       1
## renalRenal disease                               1
## endoEndocrine disease                            1
## mskMSK disease                                   1
## cancerCancer                                     1
## allergyAllergy                                   1
## psychiPsych disease                              1
## phq9_1                                           1
## phq9_2                                           1
```

```
## phq9_3                                                 1
## phq9_4                                                 1
## phq9_5                                                 1
## phq9_6                                                 1
## phq9_7                                                 1
## phq9_8                                                 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.1697e+01  on 271  degrees of freedom
## Residual deviance: 3.3399e-09  on 215  degrees of freedom
## AIC: 114
##
## Number of Fisher Scoring iterations: 25
```

```
binary_logistic_tbl <- binary_logistic_model %>% tbl_regression(exponentiate = TRUE)
binary_logistic_tbl
```

Characteristic	**OR**	**95% CI**	**p-value**
Age	1.08	0.00, Inf	>0.9
Sex			
male	—	—	
female	0.61	0.00, Inf	>0.9
BMI	1.43	0.00, Inf	>0.9
Ethnic			
Thai	—	—	
Chinese	35,435,946	0.00, Inf	>0.9
Marital status			
single	—	—	
married	0.00	0.00, Inf	>0.9
divorced	43.2	0.00, Inf	>0.9
widow	0.00	0.00, Inf	>0.9
Address			
Bangkok	—	—	
others	41.6	0.00, Inf	>0.9
Education			
not educate	—	—	
elementary	4.56	0.00, Inf	>0.9
high school	0.00	0.00, Inf	>0.9
college degree	0.33	0.00, Inf	>0.9

Characteristic	OR	95% CI	p-value
Employment			
unemployed	—	—	
part-time job	0.26	0.00, Inf	>0.9
full-time job	3,487	0.00, Inf	>0.9
retired	19.0	0.00, Inf	>0.9
income			
10,000 or less	—	—	
10,001 - 20,000	0.86	0.00, Inf	>0.9
20,001 - 30,000	5,250	0.00, Inf	>0.9
30,001 or more	872	0.00, Inf	>0.9
unknown	0.01	0.00, Inf	>0.9
Income Loss from COVID-19			
Same	—	—	
Less than 50% loss	27,602,728	0.00, Inf	>0.9
Over 50% loss	1.98	0.00, Inf	>0.9
No income	0.00	0.00, Inf	>0.9
PatientAmbulation			
Normal	—	—	
Gait aid	20,099	0.00, Inf	>0.9
Hearing			
Normal	—	—	
Hearing aid	928	0.00, Inf	>0.9
Hearing impairment	0.00	0.00, Inf	>0.9
PatientVisual			
Normal	—	—	
Glasses	0.19	0.00, Inf	>0.9
Smoking			
Never smoking	—	—	
Current smoking	19,797,474	0.00, Inf	>0.9
Past smoking	0.39	0.00, Inf	>0.9
Alcohol Drinking			
Never drinking	—	—	
Social drinking	1,980	0.00, Inf	>0.9

Characteristic	OR	95% CI	p-value
Regular drinking	8,913,392,299,128,560	0.00, Inf	>0.9
Dementia diagnosis			
No	—	—	
Yes	0.00	0.00, Inf	>0.9
Not sure	0.00	0.00, Inf	>0.9
Self Percept Cognition			
Normal	—	—	
Minor cognitive problem	347	0.00, Inf	>0.9
Major cognitive problem	1,052,635,578,977,005,936	0.00, Inf	>0.9
Number of Hospitalization	60.7	0.00, Inf	>0.9
Self Percept Health			
Worst	—	—	
Bad	0.22	0.00, Inf	>0.9
Average	7.25	0.00, Inf	>0.9
Good	171	0.00, Inf	>0.9
Best	28.4	0.00, Inf	>0.9
neuro			
None	—	—	
Neurological disease	830	0.00, Inf	>0.9
cvs			
None	—	—	
Cardiovascular disease	153	0.00, Inf	>0.9
respi			
None	—	—	
Respiratory disease	251	0.00, Inf	>0.9
gi			
None	—	—	
Gastrointestinal disease	5.67	0.00, Inf	>0.9
renal			
None	—	—	
Renal disease	0.11	0.00, Inf	>0.9
endo			
None	—	—	
Endocrine disease	11.7	0.00, Inf	>0.9
msk			

Characteristic	OR	95% CI	p-value
None	—	—	
MSK disease	0.01	0.00, Inf	>0.9
cancer			
None	—	—	
Cancer	159,438	0.00, Inf	>0.9
allergy			
None	—	—	
Allergy	0.23	0.00, Inf	>0.9
psychi			
None	—	—	
Psych disease	0.00	0.00, Inf	>0.9
Little interest or pleasure in doing things	0.63	0.00, Inf	>0.9
Feeling down, depressed, or hopeless	272	0.00, Inf	>0.9
Trouble falling or staying asleep, or sleeping too much	0.05	0.00, Inf	>0.9
Feeling tired or having little energy	195	0.00, Inf	>0.9
Poor appetite or overeating	0.38	0.00, Inf	>0.9
Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0.33	0.00, Inf	>0.9
Trouble concentrating on things, such as reading the newspaper or watching television	0.00	0.00, Inf	>0.9
Moving or speaking so slowly that other people could have noticed? Or so fidgety or restless that you have been moving a lot more than usual	2,183	0.00, Inf	>0.9

#Export data

```
table1 %>% as_hux_xlsx("../output/table_1.xlsx")
table2 %>% as_hux_xlsx("../output/table_2.xlsx")

lm_tbl %>% as_hux_xlsx("../output/multivariated_linear.xlsx")
ordered_logistic_tbl %>% as_hux_xlsx("../output/phd9_cat_ordered_logistic.xlsx")
binary_logistic_tbl %>% as_hux_xlsx("../output/phq9_9th_logistic.xlsx")
```