

```

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

```

```
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

```

```
#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
#data_filter <- readRDS(file = "../output/data_preprocessed_mergehearing.rds") # merge the hearing aid

```

```
## 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
}

```

```
##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"))
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"))
```

```
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()
```

table1

```
## Table printed with 'knitr::kable()', not {gt}. Learn why at
## https://www.danielsjoberg.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	262 (96%)
Chinese	9 (3.3%)
others	1 (0.4%)
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	

Characteristic	N = 272
unemployed	93 (34%)
part-time job	23 (8.5%)
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%)

Characteristic	N = 272
Good	131 (48%)
Best	23 (8.5%)
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%)

Characteristic	N = 272
phq9_5	
0	222 (82%)
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.danielsjoberg.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					0.008

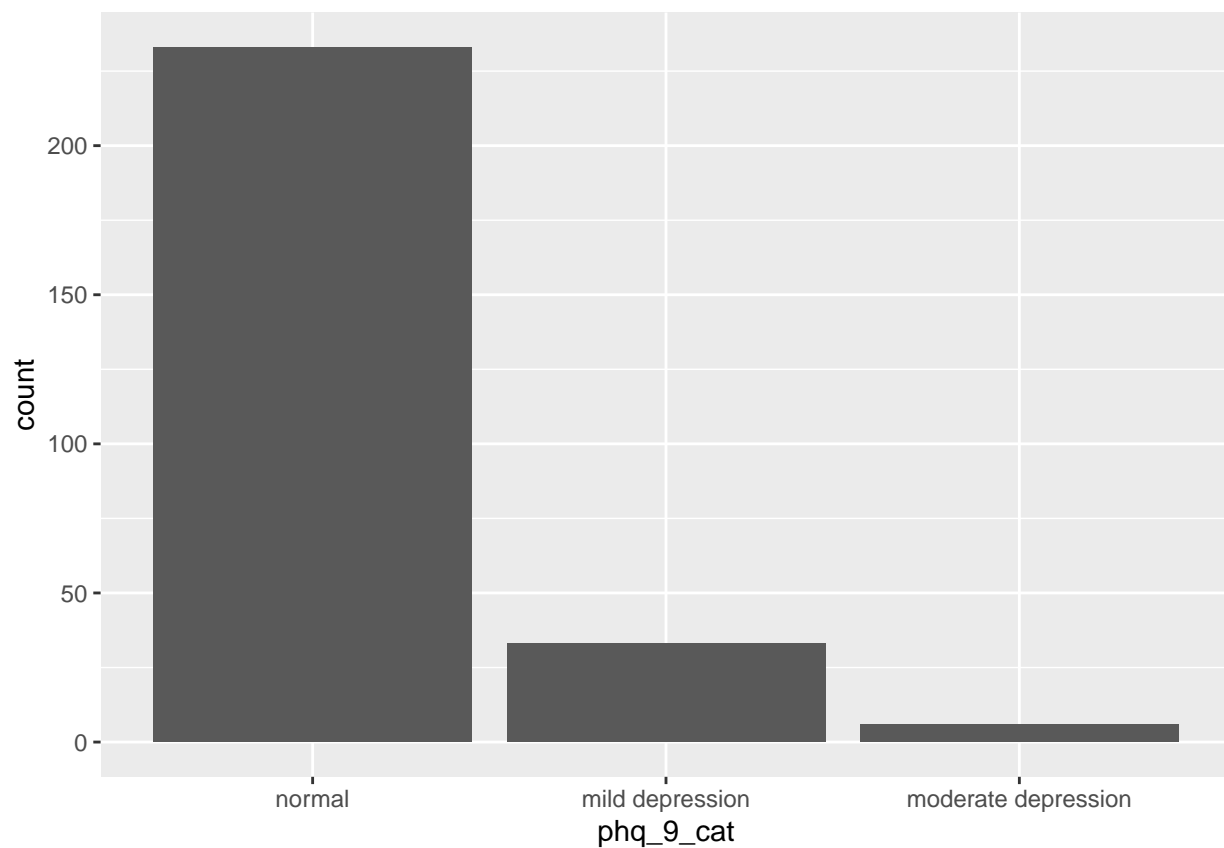
Characteristic	Overall, N = 272	normal, N = 233	mild depression, N = 33	moderate depression, N = 6	p-value
Thai	262 (96%)	226 (97%)	32 (97%)	4 (67%)	0.5
Chinese	9 (3.3%)	7 (3.0%)	1 (3.0%)	1 (17%)	
others	1 (0.4%)	0 (0%)	0 (0%)	1 (17%)	
Marital status					
single	50 (18%)	45 (19%)	5 (15%)	0 (0%)	
married	172 (63%)	149 (64%)	19 (58%)	4 (67%)	0.2
divorced	10 (3.7%)	8 (3.4%)	2 (6.1%)	0 (0%)	
widow	40 (15%)	31 (13%)	7 (21%)	2 (33%)	
Address					
Bangkok	182 (67%)	152 (65%)	24 (73%)	6 (100%)	
others	90 (33%)	81 (35%)	9 (27%)	0 (0%)	0.13
Education					
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%)	0.3
above college degree	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Employment					
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%)	0.081
retired	132 (49%)	117 (50%)	11 (33%)	4 (67%)	
income					
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%)	0.2
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					
Same	221 (81%)	194 (83%)	22 (67%)	5 (83%)	0.4
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					
Normal	260 (96%)	224 (96%)	30 (91%)	6 (100%)	0.2
Gait aid	12 (4.4%)	9 (3.9%)	3 (9.1%)	0 (0%)	
Bedbound	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Hearing					
Normal	255 (94%)	221 (95%)	29 (88%)	5 (83%)	
Hearing aid	2 (0.7%)	2 (0.9%)	0 (0%)	0 (0%)	0.4
Hearing impairment	15 (5.5%)	10 (4.3%)	4 (12%)	1 (17%)	
Visual					
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%)	0.5
Smoking					
Never smoking	234 (86%)	202 (87%)	26 (79%)	6 (100%)	

Characteristic	Overall, N = 272	normal, N = 233	mild depression, N = 33	moderate depression, N = 6	p-value
Current smoking	6 (2.2%)	5 (2.1%)	1 (3.0%)	0 (0%)	0.5
Past smoking	32 (12%)	26 (11%)	6 (18%)	0 (0%)	
Alcohol					
Drinking					
Never drinking	247 (91%)	212 (91%)	29 (88%)	6 (100%)	0.036
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					0.002
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					0.003
Cognition					
Normal	115 (42%)	107 (46%)	7 (21%)	1 (17%)	
Minor cognitive	155 (57%)	125 (54%)	26 (79%)	4 (67%)	
problem					<0.001
Major cognitive	2 (0.7%)	1 (0.4%)	0 (0%)	1 (17%)	
problem					
Number of					
Hospitalization					0.053
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%)	0.6
Self Percept					
Health					
Worst	1 (0.4%)	0 (0%)	1 (3.0%)	0 (0%)	
Bad	6 (2.2%)	2 (0.9%)	1 (3.0%)	3 (50%)	0.007
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					0.002
None	238 (88%)	208 (89%)	26 (79%)	4 (67%)	
Neurological	34 (13%)	25 (11%)	7 (21%)	2 (33%)	
disease					
cvs					0.018
None	53 (19%)	48 (21%)	4 (12%)	1 (17%)	
Cardiovascular	219 (81%)	185 (79%)	29 (88%)	5 (83%)	
disease					
respi					0.007
None	251 (92%)	218 (94%)	30 (91%)	3 (50%)	
Respiratory	21 (7.7%)	15 (6.4%)	3 (9.1%)	3 (50%)	
disease					
gi					0.002
None	216 (79%)	189 (81%)	26 (79%)	1 (17%)	
Gastrointestinal	56 (21%)	44 (19%)	7 (21%)	5 (83%)	
disease					
renal					0.018
None	250 (92%)	218 (94%)	28 (85%)	4 (67%)	
Renal disease	22 (8.1%)	15 (6.4%)	5 (15%)	2 (33%)	

Characteristic	Overall, N = 272	normal, N = 233	mild depression, N = 33	moderate depression, N = 6	p-value
endo					0.3
None	211 (78%)	184 (79%)	22 (67%)	5 (83%)	
Endocrine disease	61 (22%)	49 (21%)	11 (33%)	1 (17%)	
msk					0.018
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					0.13
None	197 (72%)	171 (73%)	24 (73%)	2 (33%)	
Allergy	75 (28%)	62 (27%)	9 (27%)	4 (67%)	
psychi					0.003
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%)	

Characteristic	Overall, N = 272	normal, N = 233	mild depression, N = 33	moderate depression, N = 6	p-value
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%)	
3	1 (0.4%)	0 (0%)	0 (0%)	1 (17%)	
phq9_9					0.007
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()
```



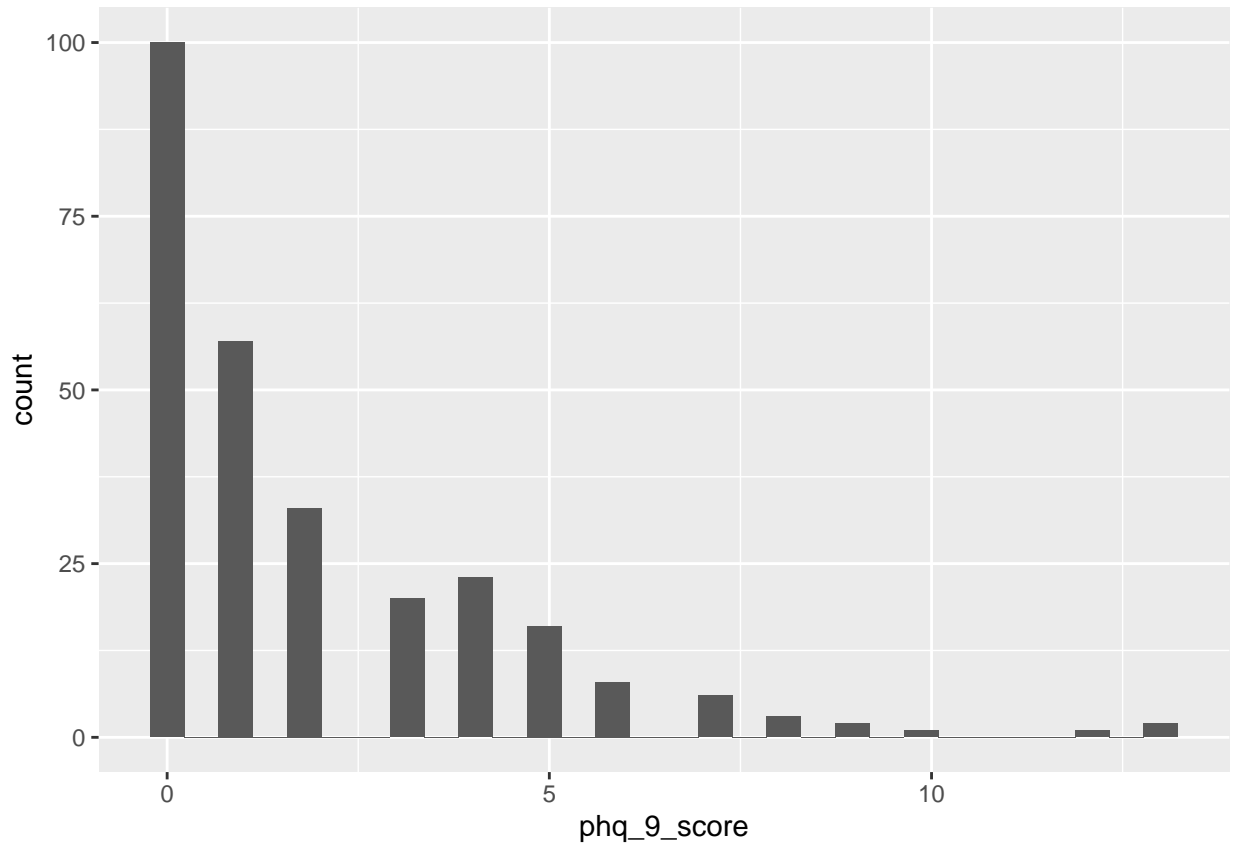
```
data_filter %>% group_by(phq_9_cat) %>%
  summarise(frequency = n())
```

```
## # 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()
```

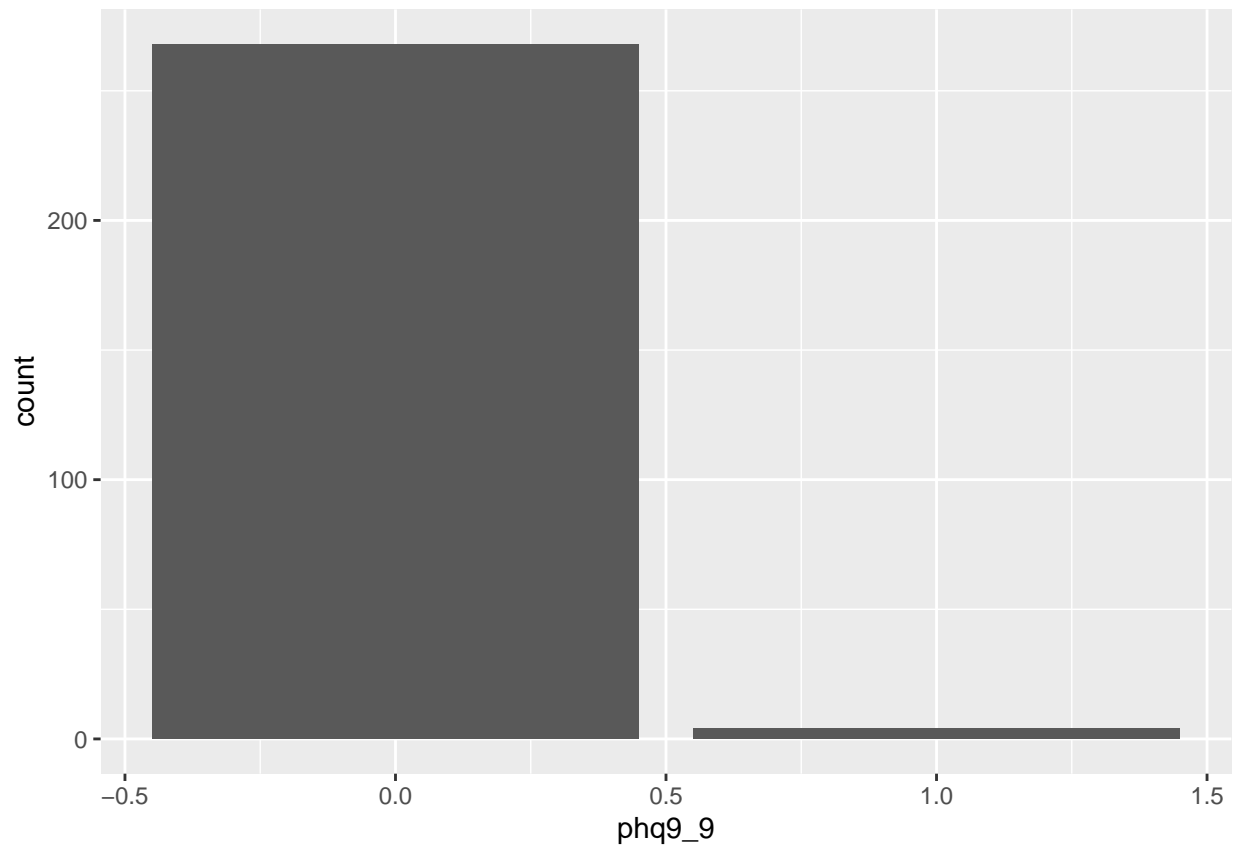
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
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
```

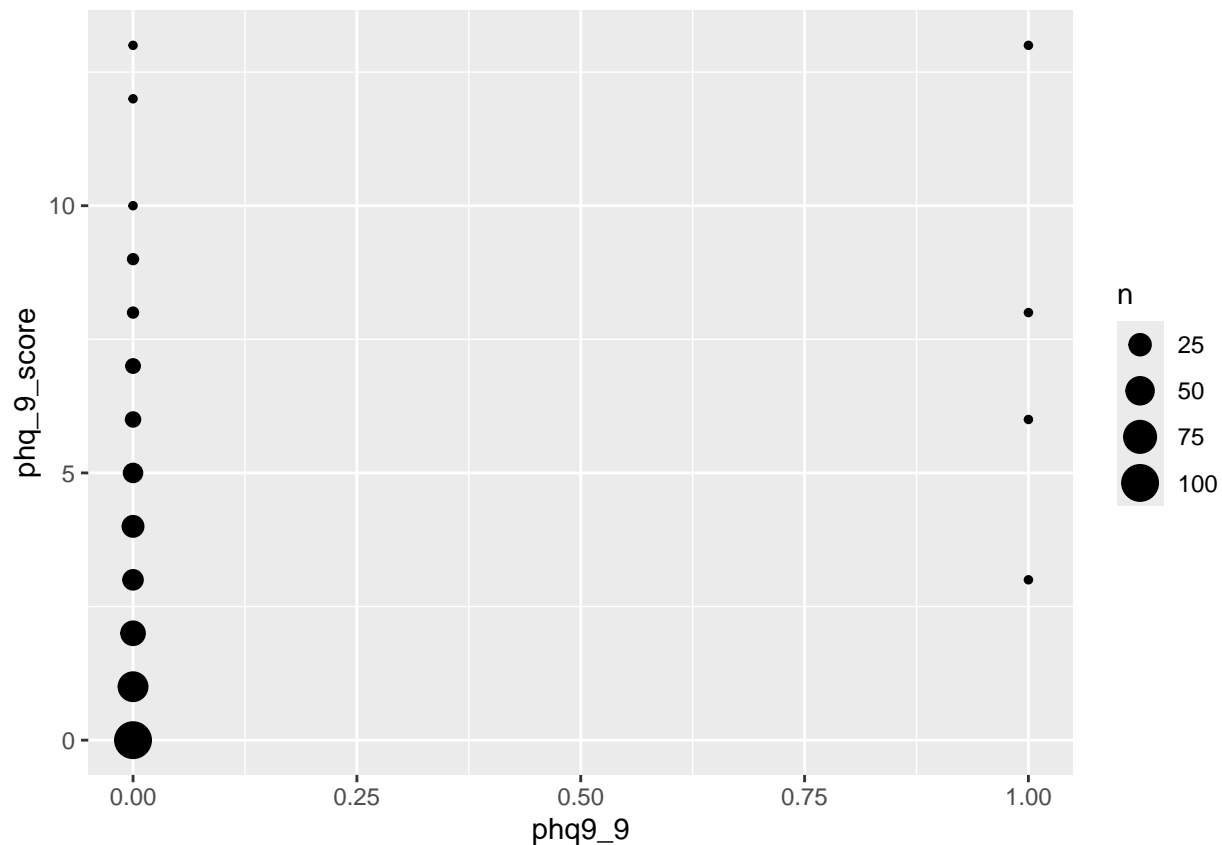
```
data_filter %>% ggplot(aes(x = phq9_9)) + geom_bar()
```



```
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 = phq9_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_score)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: data_filter[data_filter$phq9_9 == 1,]$phq_9_score and data_filter[data_filter$phq9_9 == 0,]$phq_9_score
```

```
## 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 fam
## [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 res
## [1] "PHQ9_9: Thoughts that you would be better off dead, or thoughts of hurting yourself in some way"
```

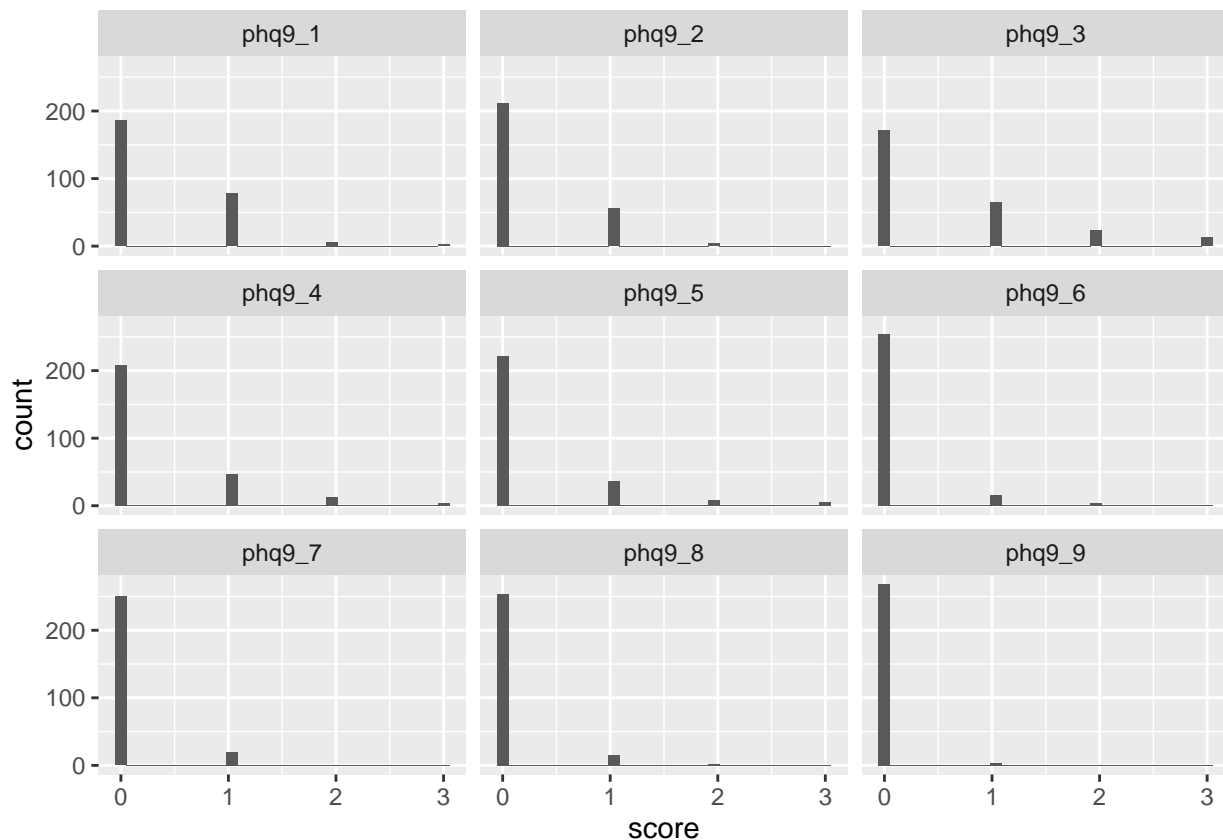
```
#convert wide to long to be used in ggplot
```

```
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 analysis
```

```
#Check for Multicollinearity
```

```
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])
```

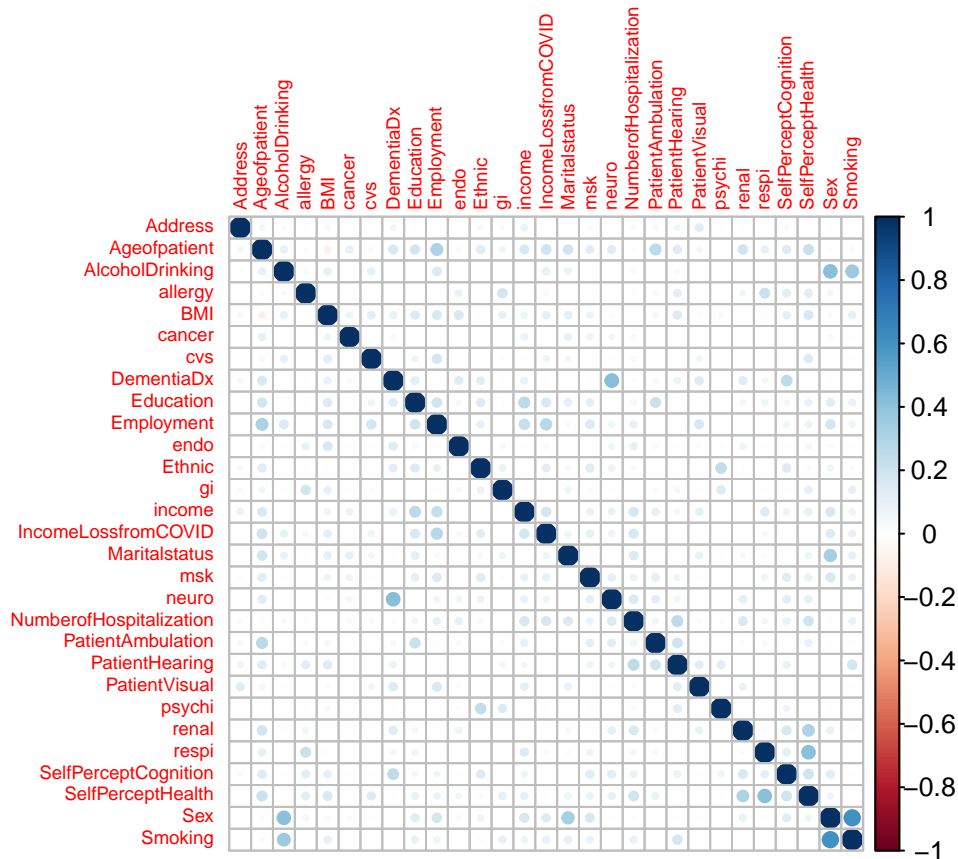
```
mix_assoc_result %>%
  dplyr::select(-complete_obs_pairs, -complete_obs_ratio) %>%
  filter(x != y) %>%
  filter(assoc != 0) %>%
  arrange(desc(abs(assoc))) %>%
  filter(row_number() %% 2 == 1) %>%
  group_by(type) %>%
  slice_max(order_by = abs(assoc), n = 10) %>%
  ungroup()
```

```
## # 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
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
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	1.637219	1	1.279539
##	Sex	2.540083	1	1.593764
##	BMI	1.334938	1	1.155395
##	Ethnic	1.737871	2	1.148165
##	Maritalstatus	1.994138	3	1.121913
##	Address	1.197663	1	1.094378
##	Education	2.784066	3	1.186078
##	Employment	3.154093	3	1.211004
##	income	3.515370	4	1.170163
##	IncomeLossfromCOVID	2.560760	3	1.169665
##	PatientAmbulation	1.386727	1	1.177594
##	PatientHearing	1.833785	2	1.163689
##	PatientVisual	1.320579	1	1.149164
##	Smoking	2.841201	2	1.298301
##	AlcoholDrinking	1.872193	2	1.169735

## DementiaDx	2.067794	2	1.199159
## SelfPerceptCognition	1.956288	2	1.182655
## NumberofHospitalization	1.407347	1	1.186316
## SelfPerceptHealth	3.032250	4	1.148737
## neuro	1.495413	1	1.222871
## cvs	1.244242	1	1.115456
## respi	1.447251	1	1.203018
## gi	1.270206	1	1.127034
## renal	1.356043	1	1.164493
## endo	1.255757	1	1.120606
## msk	1.344167	1	1.159382
## cancer	1.191778	1	1.091686
## allergy	1.292058	1	1.136687
## psychi	1.241485	1	1.114220

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)
```

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

```
selected_predictors_with_preference
```

```
## [1] "SelfPerceptHealth"      "SelfPerceptCognition"
## [3] "Ethnic"                 "NumberofHospitalization"
## [5] "gi"                     "renal"
## [7] "msk"                    "neuro"
## [9] "respi"                  "psychi"
## [11] "DementiaDx"             "allergy"
## [13] "PatientHearing"         "income"
## [15] "Ageofpatient"           "IncomeLossfromCOVID"
## [17] "Education"              "Maritalstatus"
## [19] "Employment"             "endo"
## [21] "Address"                "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] "Ethnic"                 "NumberofHospitalization"
## [5] "gi"                     "renal"
## [7] "msk"                    "neuro"
```

```

## [9] "respi"           "psychi"
## [11] "DementiaDx"      "allergy"
## [13] "PatientHearing"  "income"
## [15] "Ageofpatient"    "IncomeLossfromCOVID"
## [17] "Education"       "Maritalstatus"
## [19] "Employment"      "endo"
## [21] "Address"         "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_
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])

```

```
# 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
## -5.9781 -1.0998 -0.2634  0.7762  6.0653
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                      0.653003   3.153486   0.207
## Ageofpatient                      0.008757   0.020274   0.432
## Sexfemale                        0.509370   0.417695   1.219
## BMI                              0.022586   0.036741   0.615
## EthnicChinese                    -0.102676   0.773414  -0.133
## Ethnicothers                      9.688664   2.212492   4.379
## Maritalstatusmarried              0.180059   0.356138   0.506
## Maritalstatusdivorced             1.000456   0.733307   1.364
## Maritalstatuswidow                0.174952   0.460412   0.380
## Addressothers                     -0.384107   0.275673  -1.393
## Educationelementary               0.652114   1.294366   0.504
## Educationhigh school              0.990985   1.293269   0.766
## Educationcollege degree           0.803314   1.291998   0.622
## Employmentpart-time job           0.001460   0.571054   0.003
## Employmentfull-time job          -0.205141   0.550458  -0.373
## Employmentretired                 0.049949   0.318130   0.157
## income10,001 - 20,000             -0.092287   0.473171  -0.195
## income20,001 - 30,000             -0.572028   0.511890  -1.117
## income30,001 or more              -0.347564   0.483926  -0.718
## incomeunknown                    -0.755063   0.449771  -1.679
## IncomeLossfromCOVIDLess than 50% loss  0.581650   0.542591   1.072
## IncomeLossfromCOVIDOver 50% loss    0.437719   0.579338   0.756
## IncomeLossfromCOVIDNo income        0.321979   0.649148   0.496
## PatientAmbulationGait aid          -0.211167   0.679676  -0.311
## PatientHearingHearing aid          -0.337606   1.680003  -0.201
## PatientHearingHearing impairment    1.111425   0.582934   1.907
## PatientVisualGlasses              -0.217314   0.282489  -0.769
## SmokingCurrent smoking             -0.276589   1.012460  -0.273
## SmokingPast smoking                0.411596   0.514920   0.799
## AlcoholDrinkingSocial drinking     0.428684   0.535744   0.800
## AlcoholDrinkingRegular drinking    1.780662   1.271182   1.401
## DementiaDxYes                     0.462251   0.777142   0.595
## DementiaDxNot sure                 1.037093   0.888270   1.168
## SelfPerceptCognitionMinor cognitive problem 0.331327   0.276919   1.196
## SelfPerceptCognitionMajor cognitive problem 2.489577   1.689941   1.473
## NumberofHospitalization            0.732600   0.291157   2.516
## SelfPerceptHealthBad                1.859227   2.370633   0.784
## SelfPerceptHealthAverage           -1.691443   2.107558  -0.803
## SelfPerceptHealthGood              -2.319615   2.100267  -1.104
```

## SelfPerceptHealthBest	-2.658490	2.145089	-1.239
## neuroNeurological disease	0.463032	0.438265	1.057
## cvsCardiovascular disease	-0.247951	0.333792	-0.743
## respiRespiratory disease	0.223616	0.534207	0.419
## giGastrointestinal disease	0.757866	0.330370	2.294
## renalRenal disease	0.545760	0.506220	1.078
## endoEndocrine disease	0.261535	0.318442	0.821
## mskMSK disease	0.807756	0.285001	2.834
## cancerCancer	0.481730	0.465070	1.036
## allergyAllergy	0.589468	0.301482	1.955
## psychiPsych disease	1.049651	0.578552	1.814
##	Pr(> t)		
## (Intercept)	0.83614		
## Ageofpatient	0.66619		
## Sexfemale	0.22396		
## BMI	0.53936		
## EthnicChinese	0.89451		
## Ethnicothers	1.84e-05	***	
## Maritalstatusmarried	0.61365		
## Maritalstatusdivorced	0.17385		
## Maritalstatuswidow	0.70432		
## Addressothers	0.16491		
## Educationelementary	0.61489		
## Educationhigh school	0.44433		
## Educationcollege degree	0.53474		
## Employmentpart-time job	0.99796		
## Employmentfull-time job	0.70975		
## Employmentretired	0.87538		
## income10,001 - 20,000	0.84554		
## income20,001 - 30,000	0.26500		
## income30,001 or more	0.47338		
## incomeunknown	0.09460	.	
## IncomeLossfromCOVIDLess than 50% loss	0.28489		
## IncomeLossfromCOVIDOver 50% loss	0.45072		
## IncomeLossfromCOVIDNo income	0.62038		
## PatientAmbulationGait aid	0.75633		
## PatientHearingHearing aid	0.84092		
## PatientHearingHearing impairment	0.05786	.	
## PatientVisualGlasses	0.44254		
## SmokingCurrent smoking	0.78497		
## SmokingPast smoking	0.42495		
## AlcoholDrinkingSocial drinking	0.42447		
## AlcoholDrinkingRegular drinking	0.16267		
## DementiaDxYes	0.55258		
## DementiaDxNot sure	0.24424		
## SelfPerceptCognitionMinor cognitive problem	0.23279		
## SelfPerceptCognitionMajor cognitive problem	0.14212		
## NumberofHospitalization	0.01257	*	
## SelfPerceptHealthBad	0.43372		
## SelfPerceptHealthAverage	0.42309		
## SelfPerceptHealthGood	0.27060		
## SelfPerceptHealthBest	0.21653		
## neuroNeurological disease	0.29188		
## cvsCardiovascular disease	0.45837		

```
## respiRespiratory disease          0.67592
## giGastrointestinal disease         0.02273 *
## renalRenal disease                0.28215
## endoEndocrine disease              0.41236
## mskMSK disease                     0.00502 **
## cancerCancer                      0.30141
## allergyAllergy                     0.05181 .
## psychiPsych disease                0.07099 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.955 on 222 degrees of freedom
## Multiple R-squared:  0.4707, Adjusted R-squared:  0.3539
## F-statistic: 4.029 on 49 and 222 DF,  p-value: 5.925e-13
```

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

Characteristic	Beta	95% CI	p-value
Age	0.01	-0.03, 0.05	0.7
Sex			
male	—	—	
female	0.51	-0.31, 1.3	0.2
BMI	0.02	-0.05, 0.09	0.5
Ethnic			
Thai	—	—	
Chinese	-0.10	-1.6, 1.4	0.9
others	9.7	5.3, 14	<0.001
Marital status			
single	—	—	
married	0.18	-0.52, 0.88	0.6
divorced	1.0	-0.44, 2.4	0.2
widow	0.17	-0.73, 1.1	0.7
Address			
Bangkok	—	—	
others	-0.38	-0.93, 0.16	0.2
Education			
not educate	—	—	
elementary	0.65	-1.9, 3.2	0.6
high school	0.99	-1.6, 3.5	0.4
college degree	0.80	-1.7, 3.3	0.5
Employment			
unemployed	—	—	
part-time job	0.00	-1.1, 1.1	>0.9
full-time job	-0.21	-1.3, 0.88	0.7
retired	0.05	-0.58, 0.68	0.9
income			
10,000 or less	—	—	
10,001 - 20,000	-0.09	-1.0, 0.84	0.8
20,001 - 30,000	-0.57	-1.6, 0.44	0.3
30,001 or more	-0.35	-1.3, 0.61	0.5
unknown	-0.76	-1.6, 0.13	0.095

Characteristic	Beta	95% CI	p-value
Income Loss from COVID-19			
Same	—	—	
Less than 50% loss	0.58	-0.49, 1.7	0.3
Over 50% loss	0.44	-0.70, 1.6	0.5
No income	0.32	-0.96, 1.6	0.6
Patient Ambulation			
Normal	—	—	
Gait aid	-0.21	-1.6, 1.1	0.8
Hearing			
Normal	—	—	
Hearing aid	-0.34	-3.6, 3.0	0.8
Hearing impairment	1.1	-0.04, 2.3	0.058
Patient Visual			
Normal	—	—	
Glasses	-0.22	-0.77, 0.34	0.4
Smoking			
Never smoking	—	—	
Current smoking	-0.28	-2.3, 1.7	0.8
Past smoking	0.41	-0.60, 1.4	0.4
Alcohol Drinking			
Never drinking	—	—	
Social drinking	0.43	-0.63, 1.5	0.4
Regular drinking	1.8	-0.72, 4.3	0.2
Dementia diagnosis			
No	—	—	
Yes	0.46	-1.1, 2.0	0.6
Not sure	1.0	-0.71, 2.8	0.2
Self Percept Cognition			
Normal	—	—	
Minor cognitive problem	0.33	-0.21, 0.88	0.2
Major cognitive problem	2.5	-0.84, 5.8	0.14
Number of Hospitalization	0.73	0.16, 1.3	0.013
Self Percept Health			
Worst	—	—	
Bad	1.9	-2.8, 6.5	0.4
Average	-1.7	-5.8, 2.5	0.4
Good	-2.3	-6.5, 1.8	0.3
Best	-2.7	-6.9, 1.6	0.2
neuro			
None	—	—	
Neurological disease	0.46	-0.40, 1.3	0.3
cvs			
None	—	—	
Cardiovascular disease	-0.25	-0.91, 0.41	0.5
respi			
None	—	—	
Respiratory disease	0.22	-0.83, 1.3	0.7
gi			
None	—	—	
Gastrointestinal disease	0.76	0.11, 1.4	0.023
renal			
None	—	—	

Characteristic	Beta	95% CI	p-value
Renal disease	0.55	-0.45, 1.5	0.3
endo	—	—	
None	—	—	
Endocrine disease	0.26	-0.37, 0.89	0.4
msk	—	—	
None	—	—	
MSK disease	0.81	0.25, 1.4	0.005
cancer	—	—	
None	—	—	
Cancer	0.48	-0.43, 1.4	0.3
allergy	—	—	
None	—	—	
Allergy	0.59	0.00, 1.2	0.052
psychi	—	—	
None	—	—	
Psych disease	1.0	-0.09, 2.2	0.071

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_excludes])

## 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    0.054445  4.133e-02  1.317e+00
## Sexfemale       0.003525  9.402e-01  3.750e-03
## BMI             0.044578  7.058e-02  6.316e-01
## EthnicChinese  -1.151566  1.366e+00 -8.431e-01
## Ethnicothers   17.243228  7.234e-06  2.384e+06
## Maritalstatusmarried -0.294002  6.878e-01 -4.275e-01
## Maritalstatusdivorced  1.096362  1.215e+00  9.023e-01
## Maritalstatuswidow   0.080706  7.960e-01  1.014e-01
## Addressothers  -1.199819  6.228e-01 -1.927e+00
```

```

## Educationelementary          -0.323698  1.742e+00 -1.859e-01
## Educationhigh school         -0.335702  1.618e+00 -2.075e-01
## Educationcollege degree      -0.397623  1.625e+00 -2.447e-01
## Employmentpart-time job      -0.418845  1.039e+00 -4.032e-01
## Employmentfull-time job       0.024463  1.108e+00  2.208e-02
## Employmentretired            -0.240764  5.752e-01 -4.186e-01
## income10,001 - 20,000        -0.104294  8.065e-01 -1.293e-01
## income20,001 - 30,000        -0.775483  1.024e+00 -7.571e-01
## income30,001 or more         -0.836901  9.156e-01 -9.141e-01
## incomeunknown                -0.251850  7.280e-01 -3.459e-01
## IncomeLossfromCOVIDLess than 50% loss  1.135049  8.342e-01  1.361e+00
## IncomeLossfromCOVIDOver 50% loss  0.307896  9.553e-01  3.223e-01
## IncomeLossfromCOVIDNo income  0.875510  1.096e+00  7.985e-01
## PatientAmbulationGait aid     -0.926939  1.030e+00 -8.996e-01
## PatientHearingHearing aid    -14.610442  8.232e-07 -1.775e+07
## PatientHearingHearing impairment  1.396016  8.965e-01  1.557e+00
## PatientVisualGlasses         -0.582491  5.629e-01 -1.035e+00
## SmokingCurrent smoking       -0.095979  1.788e+00 -5.367e-02
## SmokingPast smoking          0.779104  1.035e+00  7.525e-01
## AlcoholDrinkingSocial drinking  0.485955  9.725e-01  4.997e-01
## AlcoholDrinkingRegular drinking  2.019607  1.635e+00  1.236e+00
## DementiaDxYes                0.226507  1.071e+00  2.114e-01
## DementiaDxNot sure           1.926088  1.137e+00  1.694e+00
## SelfPerceptCognitionMinor cognitive problem  0.828320  5.988e-01  1.383e+00
## SelfPerceptCognitionMajor cognitive problem  2.037383  2.150e+00  9.475e-01
## NumberofHospitalization       0.803641  4.143e-01  1.940e+00
## SelfPerceptHealthBad          -2.536437  2.879e+00 -8.810e-01
## SelfPerceptHealthAverage      -4.166473  2.398e+00 -1.738e+00
## SelfPerceptHealthGood         -4.968574  2.382e+00 -2.086e+00
## SelfPerceptHealthBest        -5.004167  2.640e+00 -1.896e+00
## neuroNeurological disease     0.382537  6.648e-01  5.754e-01
## cvsCardiovascular disease     0.308198  6.671e-01  4.620e-01
## respiRespiratory disease      0.668874  8.961e-01  7.464e-01
## giGastrointestinal disease    0.612006  5.657e-01  1.082e+00
## renalRenal disease            0.466944  7.696e-01  6.067e-01
## endoEndocrine disease         -0.219586  5.892e-01 -3.727e-01
## mskMSK disease                1.153507  5.311e-01  2.172e+00
## cancerCancer                 0.012842  7.773e-01  1.652e-02
## allergyAllergy               -0.048903  5.827e-01 -8.393e-02
## psychiPsych disease           1.597077  8.029e-01  1.989e+00
##
## Intercepts:
##                               Value      Std. Error    t value
## normal|mild depression        3.474800e+00  4.665400e+00  7.448000e-01
## mild depression|moderate depression  6.485200e+00  4.685600e+00  1.384100e+00
##
## Residual Deviance: 173.4206
## AIC: 275.4206

```

```

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

```


Characteristic	OR	95% CI	p-value
Age	1.06	0.97, 1.15	0.2
Sex			
male	—	—	
female	1.00	0.16, 6.40	>0.9
BMI	1.05	0.91, 1.20	0.5
Ethnic			
Thai	—	—	
Chinese	0.32	0.02, 4.67	0.4
others	30,806,241	30,805,802, 30,806,681	<0.001
Marital status			
single	—	—	
married	0.75	0.19, 2.89	0.7
divorced	2.99	0.27, 32.8	0.4
widow	1.08	0.23, 5.20	>0.9
Address			
Bangkok	—	—	
others	0.30	0.09, 1.03	0.055
Education			
not educate	—	—	
elementary	0.72	0.02, 22.4	0.9
high school	0.71	0.03, 17.3	0.8
college degree	0.67	0.03, 16.5	0.8
Employment			
unemployed	—	—	
part-time job	0.66	0.08, 5.09	0.7
full-time job	1.02	0.12, 9.09	>0.9
retired	0.79	0.25, 2.44	0.7
income			
10,000 or less	—	—	
10,001 - 20,000	0.90	0.18, 4.42	0.9
20,001 - 30,000	0.46	0.06, 3.47	0.4
30,001 or more	0.43	0.07, 2.63	0.4
unknown	0.78	0.19, 3.26	0.7
Income Loss from COVID-19			
Same	—	—	
Less than 50% loss	3.11	0.60, 16.1	0.2
Over 50% loss	1.36	0.21, 8.94	0.7
No income	2.40	0.28, 20.8	0.4
Ambulation			
Normal	—	—	
Gait aid	0.40	0.05, 3.02	0.4
Hearing			
Normal	—	—	
Hearing aid	0.00	0.00, 0.00	<0.001
Hearing impairment	4.04	0.69, 23.6	0.12
Visual			
Normal	—	—	
Glasses	0.56	0.18, 1.69	0.3
Smoking			
Never smoking	—	—	
Current smoking	0.91	0.03, 30.8	>0.9
Past smoking	2.18	0.28, 16.8	0.5

Characteristic	OR	95% CI	p-value
Alcohol Drinking			
Never drinking	—	—	
Social drinking	1.63	0.24, 11.1	0.6
Regular drinking	7.54	0.30, 189	0.2
Dementia diagnosis			
No	—	—	
Yes	1.25	0.15, 10.4	0.8
Not sure	6.86	0.73, 64.5	0.092
Self Percept Cognition			
Normal	—	—	
Minor cognitive problem	2.29	0.70, 7.45	0.2
Major cognitive problem	7.67	0.11, 531	0.3
Number of Hospitalization	2.23	0.99, 5.05	0.054
Self Percept Health			
Worst	—	—	
Bad	0.08	0.00, 23.0	0.4
Average	0.02	0.00, 1.75	0.084
Good	0.01	0.00, 0.76	0.038
Best	0.01	0.00, 1.22	0.059
neuro			
None	—	—	
Neurological disease	1.47	0.40, 5.43	0.6
cvs			
None	—	—	
Cardiovascular disease	1.36	0.37, 5.07	0.6
respi			
None	—	—	
Respiratory disease	1.95	0.33, 11.4	0.5
gi			
None	—	—	
Gastrointestinal disease	1.84	0.60, 5.62	0.3
renal			
None	—	—	
Renal disease	1.60	0.35, 7.27	0.5
endo			
None	—	—	
Endocrine disease	0.80	0.25, 2.56	0.7
msk			
None	—	—	
MSK disease	3.17	1.11, 9.03	0.031
cancer			
None	—	—	
Cancer	1.01	0.22, 4.69	>0.9
allergy			
None	—	—	
Allergy	0.95	0.30, 3.00	>0.9
psychi			
None	—	—	
Psych disease	4.94	1.01, 24.0	0.048

```
Anova(ordered_logistic_model)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: phq_9_cat
##
##          LR Chisq Df Pr(>Chisq)
## Ageofpatient      1.7876 1  0.18122
## Sex                0.0000 1  0.99728
## BMI               0.3977 1  0.52829
## Ethnic            5.3320 2  0.06953 .
## Maritalstatus     1.6888 3  0.63942
## Address           4.1379 1  0.04193 *
## Education         0.0649 4  0.99948
## Employment        0.2974 3  0.96052
## income            1.3426 4  0.85412
## IncomeLossfromCOVID 2.0402 3  0.56411
## PatientAmbulation  0.8515 2  0.65328
## PatientHearing     2.9033 2  0.23419
## PatientVisual      1.1133 2  0.57313
## Smoking            0.7155 2  0.69925
## AlcoholDrinking    1.5049 2  0.47122
## DementiaDx         2.8161 2  0.24462
## SelfPerceptCognition 2.6744 2  0.26258
## NumberofHospitalization 3.5292 1  0.06030 .
## SelfPerceptHealth  8.9286 4  0.06291 .
## neuro             0.3259 1  0.56809
## cvs               0.2177 1  0.64081
## respi            0.5343 1  0.46481
## gi               1.1475 1  0.28408
## renal            0.3568 1  0.55027
## endo             0.1414 1  0.70686
## msk              4.9475 1  0.02613 *
## cancer            0.0003 1  0.98691
## allergy           0.0070 1  0.93315
## psychi           3.7056 1  0.05423 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Binary logistic regression
```

```
#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", "phq9_9")

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) %in% column_to_excludes])

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)      -5.941e+01  7.595e+05      0
## Ageofpatient       6.918e-02  5.100e+03      0
## Sexfemale         -2.571e-01  1.406e+05      0
## BMI               3.464e-01  1.218e+04      0
## EthnicChinese     1.735e+01  2.008e+05      0
## Ethnicothers      1.624e+01  6.439e+05      0
## Maritalstatusmarried -9.738e+00  8.372e+04      0
## Maritalstatusdivorced 3.892e+00  1.426e+05      0
## Maritalstatuswidow  -1.212e+01  1.253e+05      0
## Addressothers      3.647e+00  8.541e+04      0
## Educationelementary 1.624e+00  3.818e+05      0
## Educationhigh school -6.768e+00  4.378e+05      0
## Educationcollege degree -7.657e-01  4.394e+05      0
## Employmentpart-time job -1.883e+00  2.210e+05      0
## Employmentfull-time job 8.141e+00  1.293e+05      0
## Employmentretired 2.803e+00  1.204e+05      0
## income10,001 - 20,000 -2.877e-01  1.525e+05      0
## income20,001 - 30,000 8.225e+00  1.677e+05      0
## income30,001 or more 6.614e+00  1.700e+05      0
## incomeunknown      -5.184e+00  1.635e+05      0
## IncomeLossfromCOVIDLess than 50% loss 1.723e+01  8.335e+04      0
## IncomeLossfromCOVIDOver 50% loss -5.669e-02  2.231e+05      0
## IncomeLossfromCOVIDNo income -8.915e+00  2.377e+05      0
## PatientAmbulationGait aid 1.033e+01  2.204e+05      0
## PatientHearingHearing aid 6.669e+00  4.186e+05      0
## PatientHearingHearing impairment -8.375e+00  2.826e+05      0
## PatientVisualGlasses -1.709e+00  1.037e+05      0
## SmokingCurrent smoking 1.671e+01  1.577e+05      0
## SmokingPast smoking -5.262e-01  1.705e+05      0
## AlcoholDrinkingSocial drinking 7.635e+00  1.152e+05      0
## AlcoholDrinkingRegular drinking 3.745e+01  2.079e+05      0
## DementiaDxYes      -1.213e+01  2.456e+05      0
## DementiaDxNot sure -2.571e+01  3.174e+05      0
## SelfPerceptCognitionMinor cognitive problem 5.795e+00  9.338e+04      0
## SelfPerceptCognitionMajor cognitive problem 4.183e+01  3.760e+05      0
## NumberofHospitalization 4.153e+00  8.349e+04      0
## SelfPerceptHealthBad -1.932e+00  6.729e+05      0
## SelfPerceptHealthAverage 1.195e+00  5.254e+05      0
## SelfPerceptHealthGood 4.379e+00  5.226e+05      0
## SelfPerceptHealthBest 2.902e+00  5.041e+05      0
## neuroNeurological disease 6.732e+00  9.112e+04      0
## cvsCardiovascular disease 5.248e+00  1.170e+05      0
## respiRespiratory disease 5.466e+00  1.629e+05      0
## giGastrointestinal disease 1.654e+00  9.940e+04      0
## renalRenal disease -1.760e+00  2.078e+05      0
## endoEndocrine disease 2.597e+00  1.102e+05      0
## mskMSK disease     -5.011e+00  7.886e+04      0
## cancerCancer       1.186e+01  8.723e+04      0
## allergyAllergy     -1.449e+00  7.167e+04      0
```

## psychiPsych disease	-7.496e+00	1.754e+05	0
## phq9_1	-1.872e-01	9.887e+04	0
## phq9_2	5.428e+00	8.907e+04	0
## phq9_3	-3.084e+00	6.248e+04	0
## phq9_4	5.226e+00	6.380e+04	0
## phq9_5	-1.010e+00	5.966e+04	0
## phq9_6	-1.512e+00	1.651e+05	0
## phq9_7	-9.118e+00	2.451e+05	0
## phq9_8	7.613e+00	1.442e+05	0
##	Pr(> z)		
## (Intercept)	1		
## Ageofpatient	1		
## Sexfemale	1		
## BMI	1		
## EthnicChinese	1		
## Ethnicothers	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: 3.3324e-09 on 214 degrees of freedom
## AIC: 116
##
## 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	1.07	0.00, Inf	>0.9
Sex			
male	—	—	
female	0.77	0.00, Inf	>0.9
BMI	1.41	0.00, Inf	>0.9
Ethnic			
Thai	—	—	
Chinese	34,277,944	0.00, Inf	>0.9
others	11,330,960	0.00, Inf	>0.9
Marital status			
single	—	—	
married	0.00	0.00, Inf	>0.9
divorced	49.0	0.00, Inf	>0.9
widow	0.00	0.00, Inf	>0.9
Address			
Bangkok	—	—	

Characteristic	OR	95% CI	p- value
others	38.4	0.00, Inf	>0.9
Education			
not educate	—	—	
elementary	5.07	0.00, Inf	>0.9
high school	0.00	0.00, Inf	>0.9
college degree	0.46	0.00, Inf	>0.9
Employment			
unemployed	—	—	
part-time job	0.15	0.00, Inf	>0.9
full-time job	3,433	0.00, Inf	>0.9
retired	16.5	0.00, Inf	>0.9
income			
10,000 or less	—	—	
10,001 - 20,000	0.75	0.00, Inf	>0.9
20,001 - 30,000	3,733	0.00, Inf	>0.9
30,001 or more	745	0.00, Inf	>0.9
unknown	0.01	0.00, Inf	>0.9
Income Loss from COVID-19			
Same	—	—	
Less than 50% loss	30,388,083	0.00, Inf	>0.9
Over 50% loss	0.94	0.00, Inf	>0.9
No income	0.00	0.00, Inf	>0.9
PatientAmbulation			
Normal	—	—	
Gait aid	30,767	0.00, Inf	>0.9
Hearing			
Normal	—	—	
Hearing aid	787	0.00, Inf	>0.9
Hearing impairment	0.00	0.00, Inf	>0.9
PatientVisual			
Normal	—	—	
Glasses	0.18	0.00, Inf	>0.9
Smoking			

Characteristic	OR	95% CI	p- value
Never smoking	—	—	
Current smoking	18,103,897	0.00, Inf	>0.9
Past smoking	0.59	0.00, Inf	>0.9
Alcohol Drinking			
Never drinking	—	—	
Social drinking	2,069	0.00, Inf	>0.9
Regular drinking	18,415,915,782,777,002	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	329	0.00, Inf	>0.9
Major cognitive problem	1,474,029,620,826,080,704	0.00, Inf	>0.9
Number of Hospitalization	63.7	0.00, Inf	>0.9
Self Percept Health			
Worst	—	—	
Bad	0.14	0.00, Inf	>0.9
Average	3.30	0.00, Inf	>0.9
Good	79.8	0.00, Inf	>0.9
Best	18.2	0.00, Inf	>0.9
neuro			
None	—	—	
Neurological disease	839	0.00, Inf	>0.9
cvs			
None	—	—	
Cardiovascular disease	190	0.00, Inf	>0.9
respi			
None	—	—	
Respiratory disease	237	0.00, Inf	>0.9
gi			
None	—	—	
Gastrointestinal disease	5.23	0.00, Inf	>0.9

Characteristic	OR	95% CI	p-value
renal	—	—	
None			
Renal disease	0.17	0.00, Inf	>0.9
endo	—	—	
None			
Endocrine disease	13.4	0.00, Inf	>0.9
msk	—	—	
None			
MSK disease	0.01	0.00, Inf	>0.9
cancer	—	—	
None			
Cancer	141,485	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	0.83	0.00, Inf	>0.9
Feeling down, depressed, or hopeless	228	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	186	0.00, Inf	>0.9
Poor appetite or overeating	0.36	0.00, Inf	>0.9
Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0.22	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	2,023	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")
```