Individual_project_Nane_Abrahamyan

R Markdown

** comments** As we see out data consist of 33 features and 30088 observations

The majority of which is categorical some already in factorized form. 1. In this analysis I want to find out whether being locked with many people(sharing same household) can increse stress level. 2. Secondly I checked which gender is more responsible about wearing masks and to see whether the masks are effective of not.

The point of analysis is to find what symtopms usually appear during covid, to see rates depending on dates, region and people's behavior(for example wearing masks or not)

3. Checking whether people with health issues are stressed about the virus, are they wearing masks or not. ### First finding
4. Let's start with some data cleaning and preparation

I won't be using columns guid and userAgent => I will remove them

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

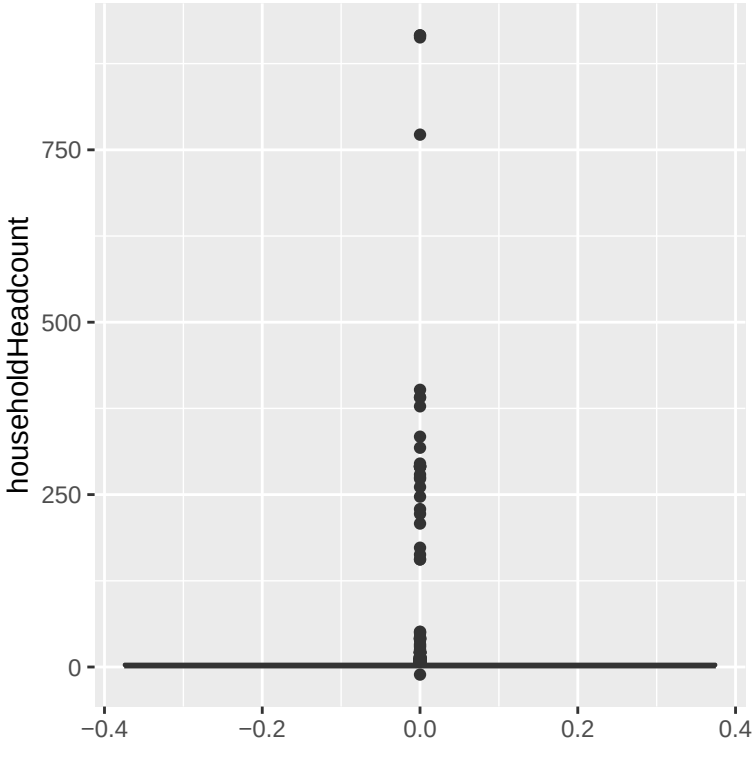
Let's start with some graph plots!!!

before ploting graphs I want to be sure that there are no outliers that can make my data biased. So I will check for outliers. Also as observed there was class imbalance problem in my data I also took equal amount from each class for having fair results.

We can check for outliers in two ways in first way we will count numbers of each household in a table and secnd one using boxplot

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 56 x 2
##    householdHeadcount Count
##                 <dbl> <int>
##  1                  2 10258
##  2                  1  6213
##  3                  4  5168
##  4                  3  5162
##  5                  5  2008
##  6                  6   684
##  7                  7   297
##  8                  8   101
##  9                  9    57
## 10                 10    32
## # ... with 46 more rows
```



So it become obvious that there are a lot of outliers in our data from which we need to get rid of. I think some users entered random numbers. There was also class imbalance problem so in following step I chose equal number of classes from householdHeadcount

```
## [1] 4.0 2.0 1.0 3.0 5.0 1.3 6.0 1.5
```

```
## -- Attaching packages ---------------------------------------------------------------------- tid
```

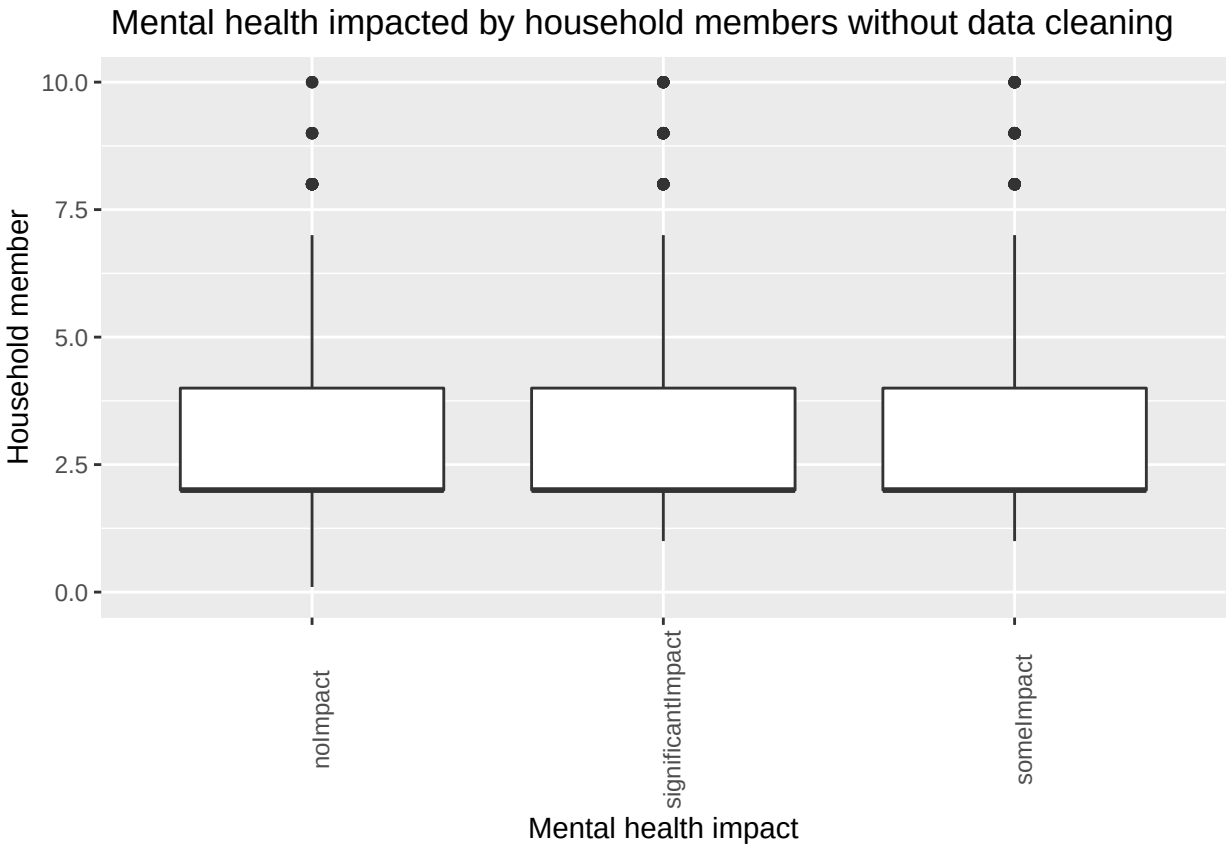
```
## v tibble  3.0.1     v purrr   0.3.4
## v tidyr   1.1.0     v forcats 0.5.0
## v readr   1.3.1
```

```
## -- Conflicts ------------------------------------------------------------------------- tidyverse
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

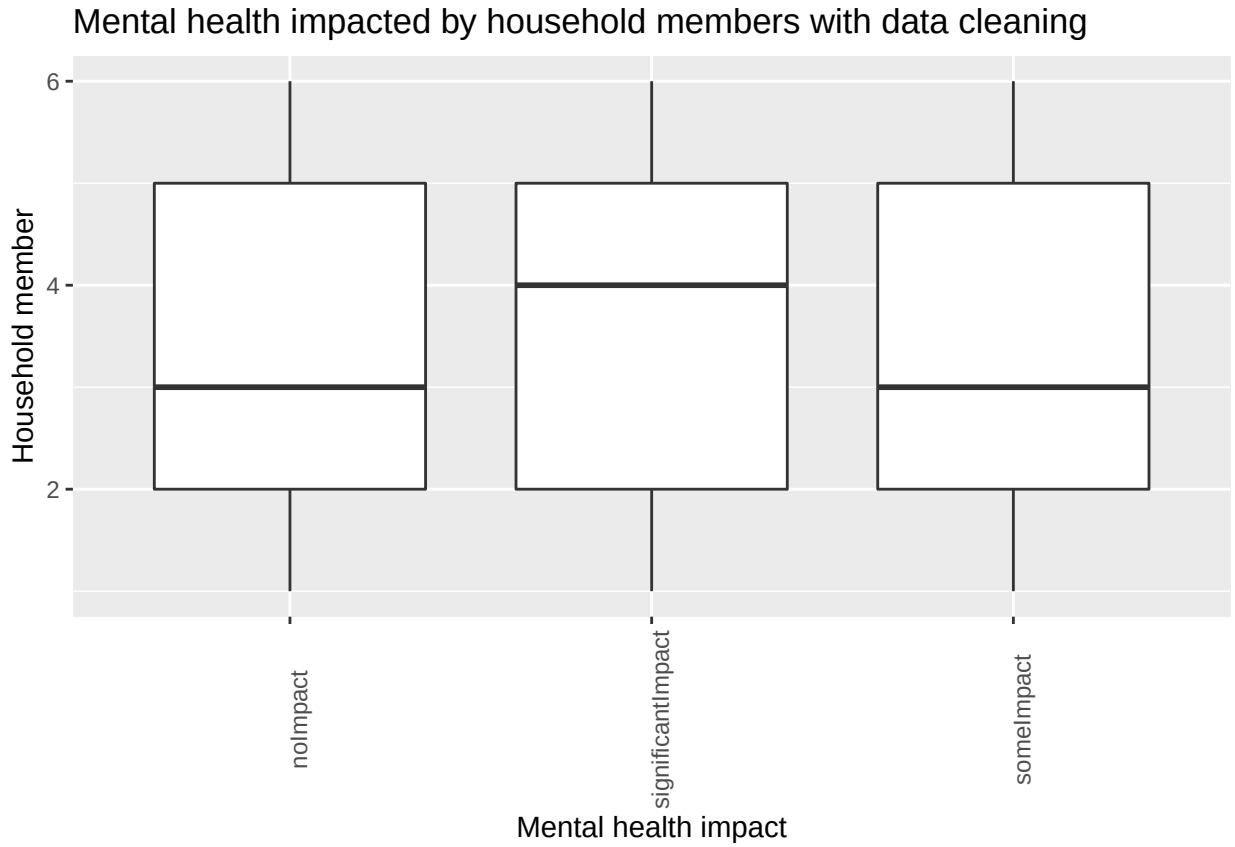
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 6 x 2
##   householdHeadcount Count
##                <dbl> <int>
## 1                  1   600
## 2                  2   600
## 3                  3   600
## 4                  4   600
## 5                  5   600
## 6                  6   600
```

let's see what can cause mental healt impact. There can be two options whether people feel stressed when there are few people at home or stressed that they are locked with many household members. As we see from the graph number of household members



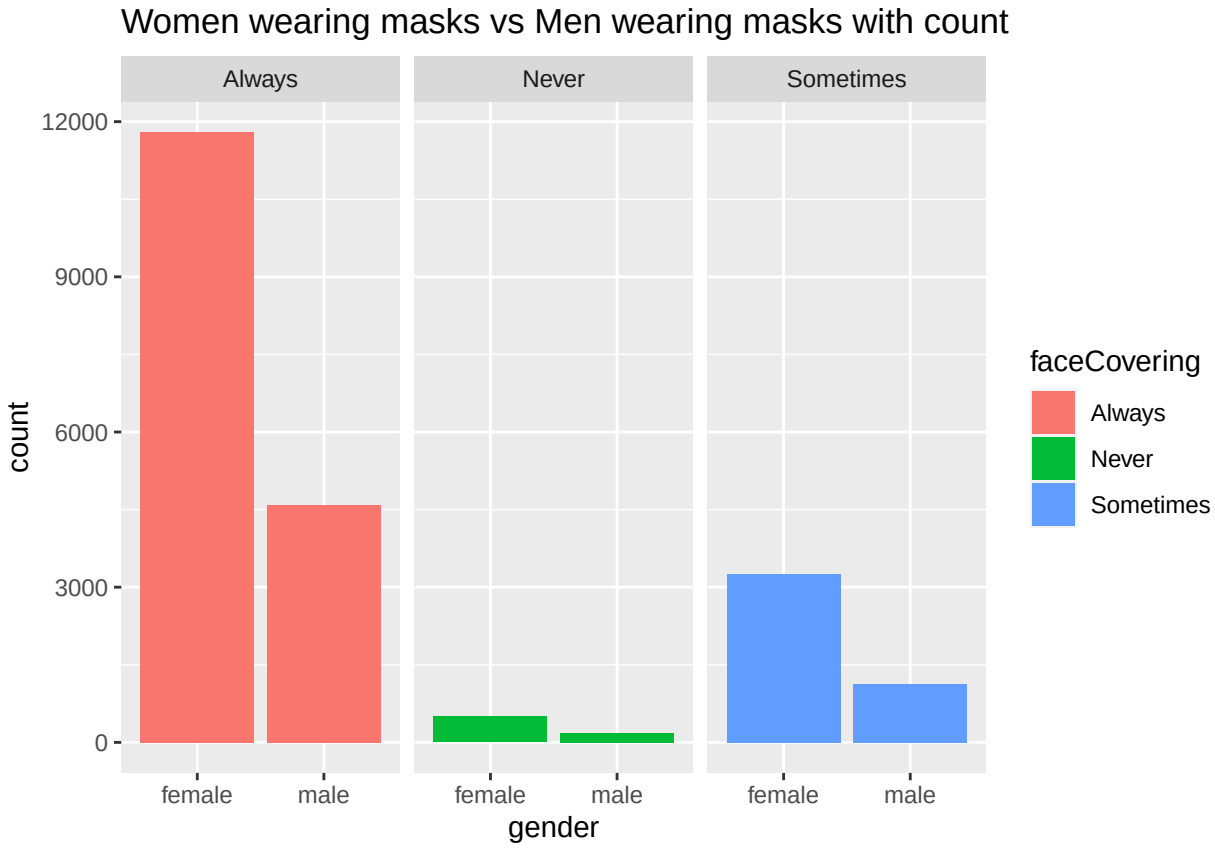

```
## [1] "someImpact"        "significantImpact" "noImpact"
```



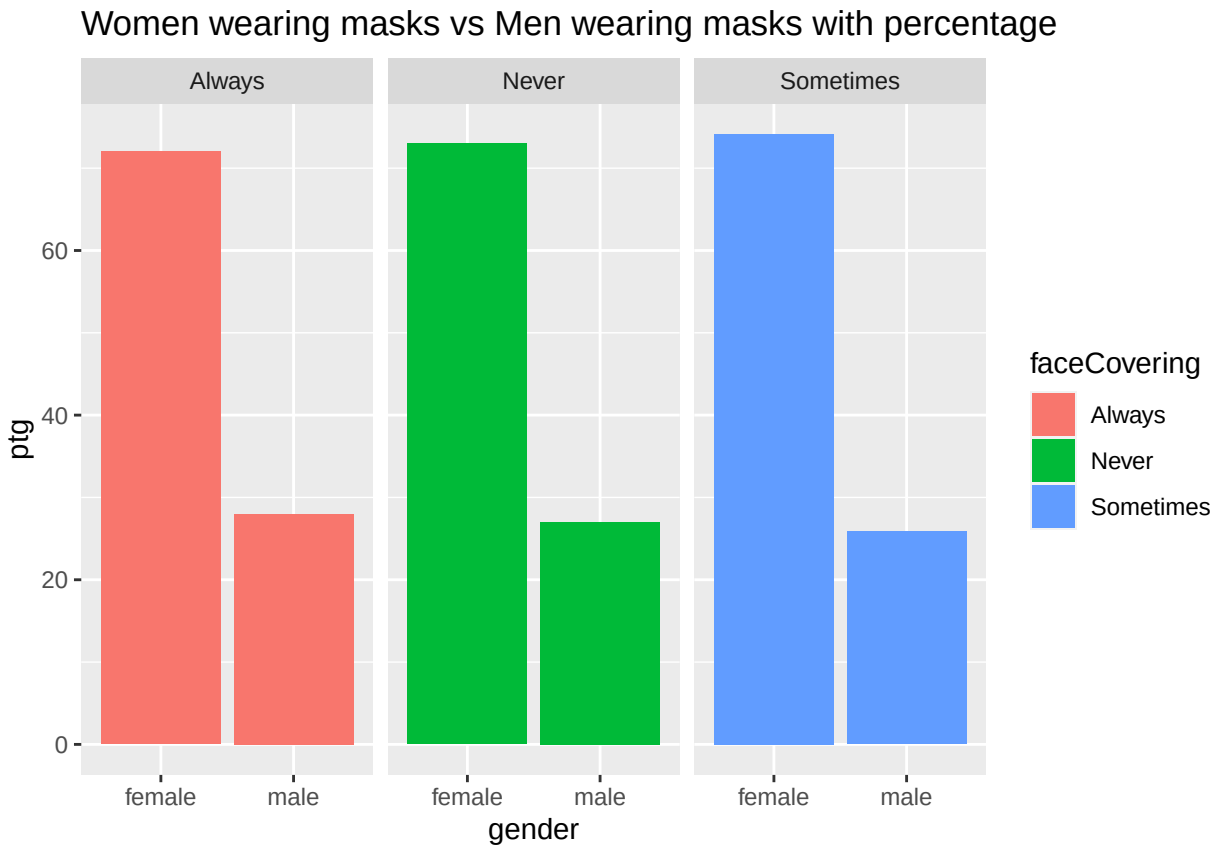
Result: the larger is number of household members the stressed people feel during the lockdown. Recommendation: sometimes leave the house :))

Second Observation

```
## `summarise()` regrouping output by 'faceCovering' (override with `.groups` argument)
```

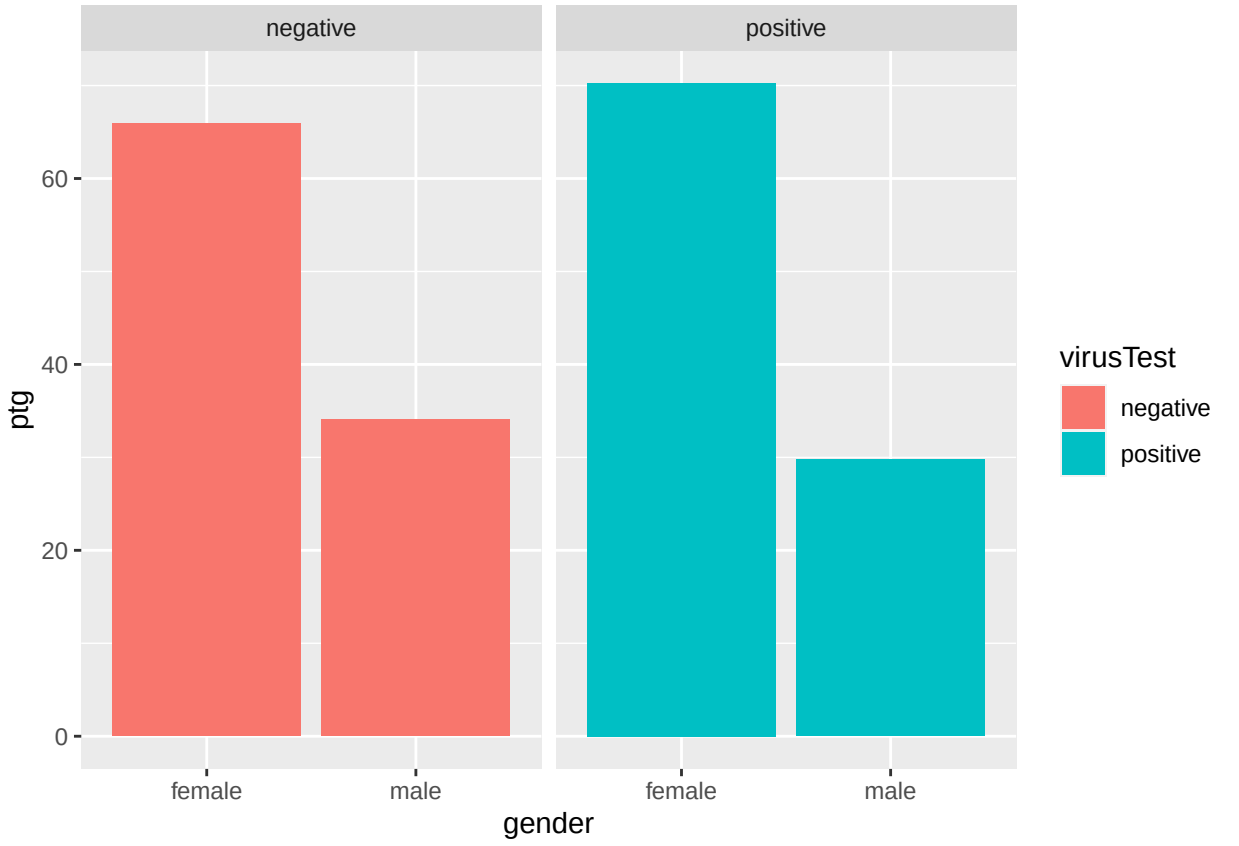



```
## `summarise()` regrouping output by 'faceCovering' (override with `.groups` argument)
```



As we see women were more responsible and wore their masks now let's see whether masks helped hell women to get infected. I think yes, but let's check.

```
## `summarise()` regrouping output by 'virusTest' (override with `.groups` argument)
```



As we masks weren't that effective as there are more percentage of infected women than men!!

3rd observation

Now let's see how people with Chronic illnesses feel about covid 1. as we see below people with chronic ilnesses are more than concerned and are stressed for their lives

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 2
##   mentalHealthImpact Count
##   <chr>              <int>
## 1 noImpact             338
## 2 significantImpact   1022
## 3 someImpact          1438
```

Below we see that people have no issues are also concerned but many said that virus didn't give them any kind of stress while there were very few people with ilnesses who chose no impact => people with illnesses are really scared

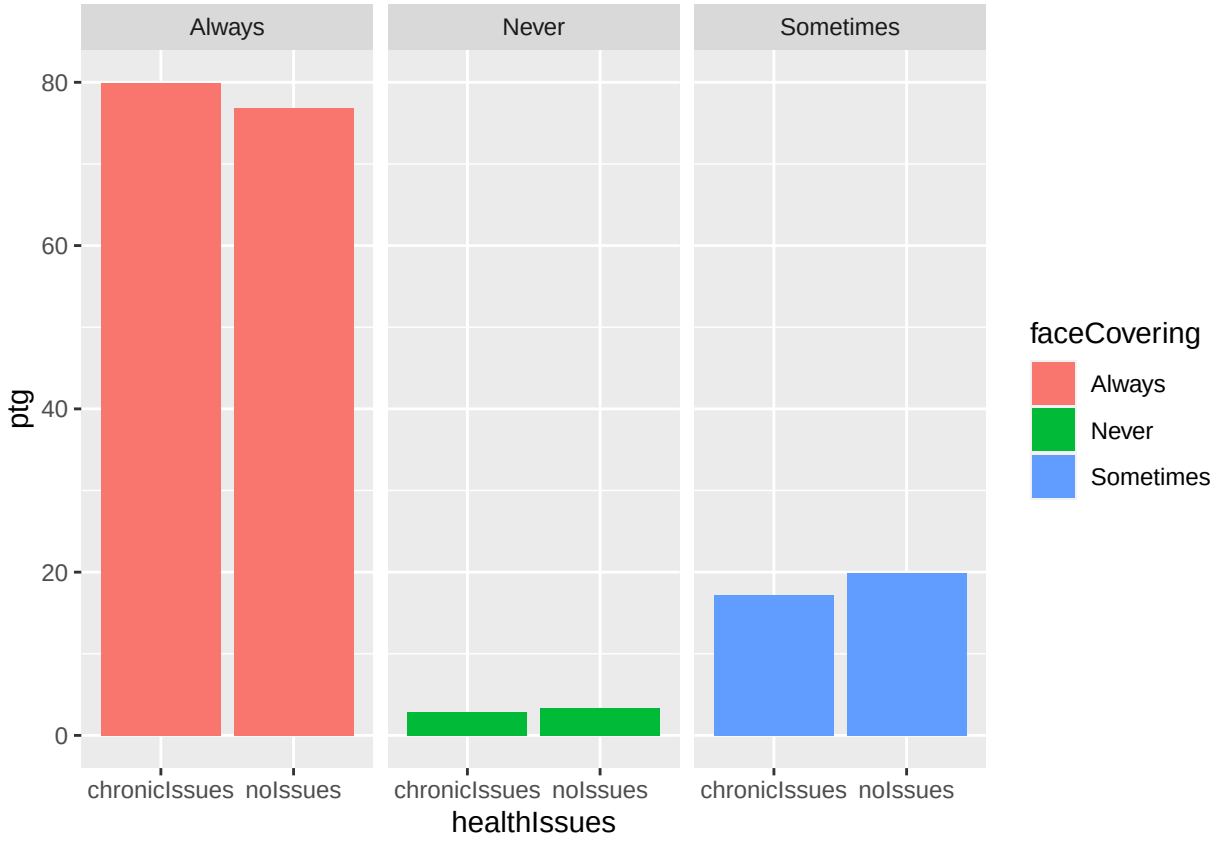
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 2
```

```
##   mentalHealthImpact Count
##   <chr>              <int>
## 1 noImpact            2840
## 2 significantImpact   3555
## 3 someImpact         10118
```

Now let's see if people with illnesses who are scared wear masks and if they do are they effective

```
## `summarise()` regrouping output by 'healthIssues' (override with `.groups` argument)
```



As expected people with illnesses wear masks more often that people with no issues

```
## `summarise()` regrouping output by 'healthIssues' (override with `.groups` argument)
```

```
## # A tibble: 4 x 3
## # Groups:   healthIssues [2]
##   healthIssues  virusTest   ptg
##   <chr>         <chr>     <dbl>
## 1 chronicIssues negative  96.0
## 2 chronicIssues positive   3.97
## 3 noIssues      negative  95.6
## 4 noIssues      positive   4.43
```

ALthough people don't get virus often people with no issues caught them more => masks were effective (the pecentage is very small but anyways it's a result)

Result: Wearing masks is effective so people should wear masks by which they would not only protect themselves but reduce stress level of people with illnesses

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.