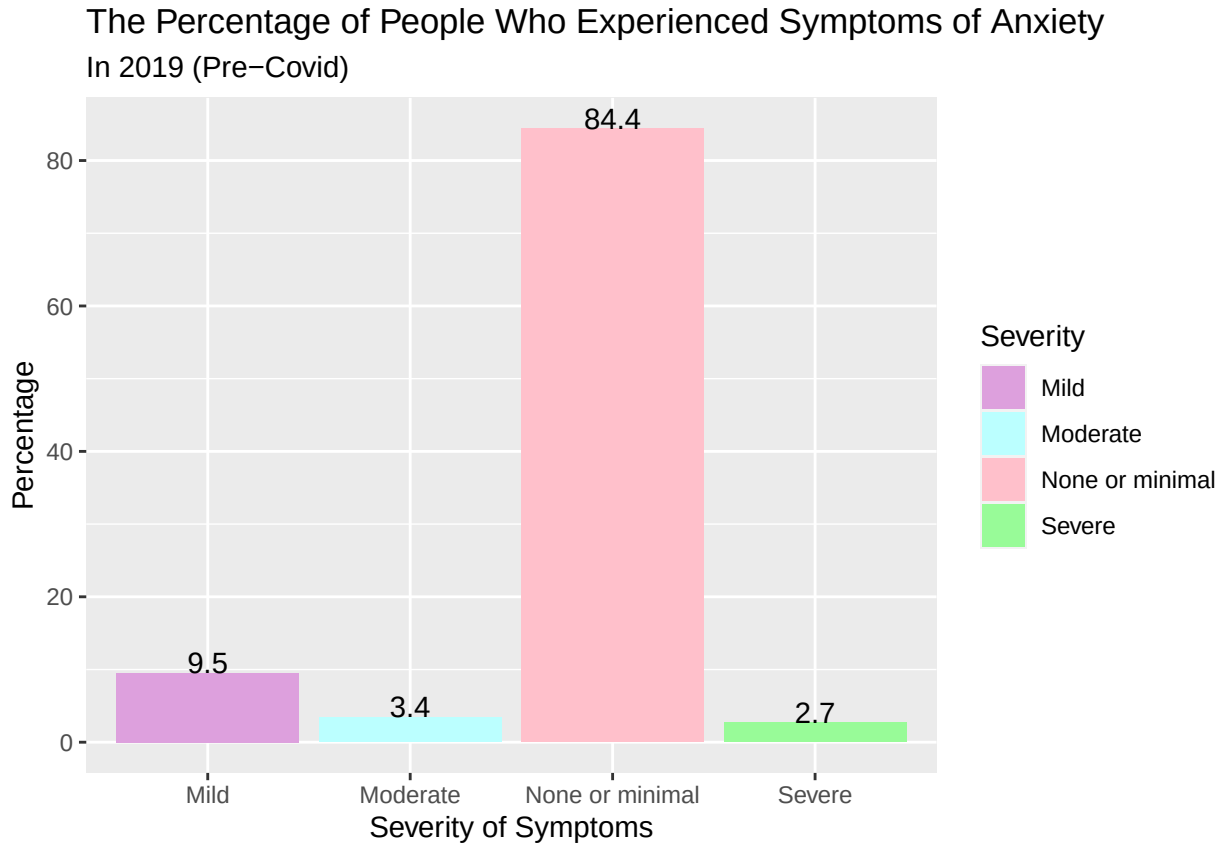
kayla

```r
# create dataset from CDC
precovid_anxiety <- data.frame(Severity = c("None or minimal", "Mild", "Moderate", "Severe"),
                               Percentage = c(84.4, 9.5, 3.4, 2.7))

# create bar chart
precovid_barchart <- ggplot(precovid_anxiety,
                            aes(x = Severity, y = Percentage, fill = Severity)) +
  geom_col() +
  geom_text(aes(label = Percentage), vjust=0) +
  scale_fill_manual("Severity",
                    values = c("Mild" = "plum",
                               "Moderate" = "paleturquoise1",
                               "None or minimal" = "pink",
                               "Severe" = "palegreen")) +
  labs(title = "The Percentage of People Who Experienced Symptoms of Anxiety",
       subtitle = "In 2019 (Pre-Covid)",
       x = "Severity of Symptoms",
       y = "Percentage")

precovid_barchart
```



How was mental health pre-covid? * During 2019, 9.5% of adults experienced mild, 3.4% experienced moderate, and 2.7% experienced severe symptoms of anxiety * 84.4% of adults aged 18 and over experienced no or minimal symptoms of anxiety

```
# import data
mental_health <- read_csv("mental_health.csv") %>%
  filter(State == "National")
```

```
## Rows: 1989 Columns: 6
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): State
## dbl (5): EST_ST, WEEK, count, anxiety_percentage, depression_percentage
```

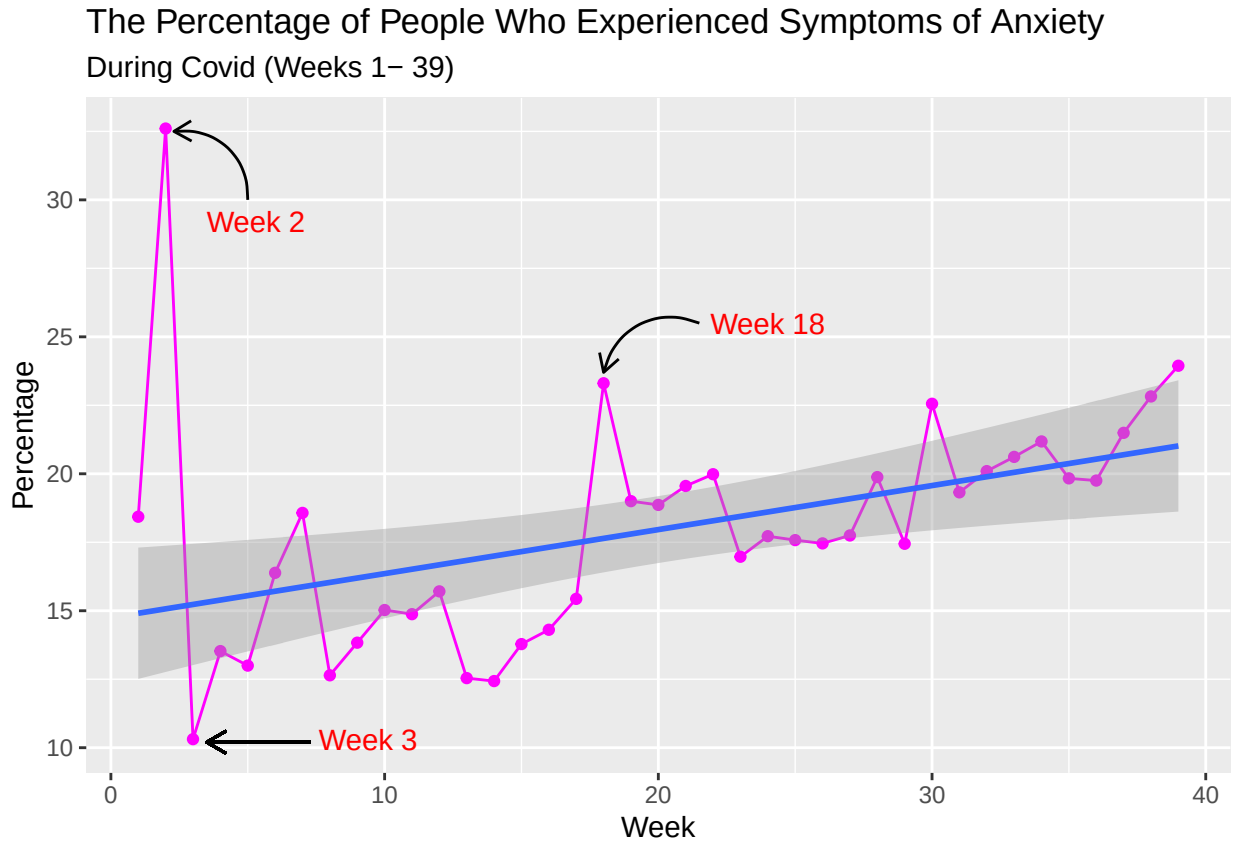
```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# plot anxiety graph with ggplot
anxiety_graph <- ggplot(data = mental_health,
           mapping = aes(x = WEEK,
                         y = anxiety_percentage)) +
      geom_line(aes(x = WEEK,
```

```
                    y = anxiety_percentage),
                color = "magenta") +
      geom_point(aes(x = WEEK,
                     y = anxiety_percentage),
                 color = "magenta") +
  geom_smooth(method='lm', formula = y~x) +
  labs(title = "The Percentage of People Who Experienced Symptoms of Anxiety",
       subtitle = "During Covid (Weeks 1- 39)",
       x = "Week",
       y = "Percentage")

df1 <- data.frame(x1 = 5, x2 = 2.3, y1 = 30, y2 = 32.5)
df2 <- data.frame(x1 = 21.5, x2 = 18, y1 = 25.5, y2 = 23.7)

anxiety_graph +
  annotate("text", x = 5.3, y = 29.2, label = "Week 2", colour = "red") +
  geom_curve(
    aes(x = x1, y = y1, xend = x2, yend = y2),
    inherit.aes = FALSE,
    data = df1,
    arrow = arrow(length = unit(0.03, "npc"))) +
  annotate("text", x = 9.4, y = 10.3, label = "Week 3", colour = "red") +
  geom_segment(aes(x = 7.3, y = 10.2, xend = 3.5, yend = 10.2),
               arrow = arrow(length = unit(0.3, "cm"))) +
  annotate("text", x = 24, y = 25.5, label = "Week 18", colour = "red") +
  geom_curve(
    aes(x = x1, y = y1, xend = x2, yend = y2),
    inherit.aes = FALSE,
    data = df2,
    arrow = arrow(length = unit(0.03, "npc")))
```




```
# using mosaic to find statistics and conditions
cor(mental_health$anxiety_percentage, mental_health$WEEK)
```

```
## [1] 0.4424454
```

```
mean(mental_health$anxiety_percentage)
```

```
## [1] 17.96227
```

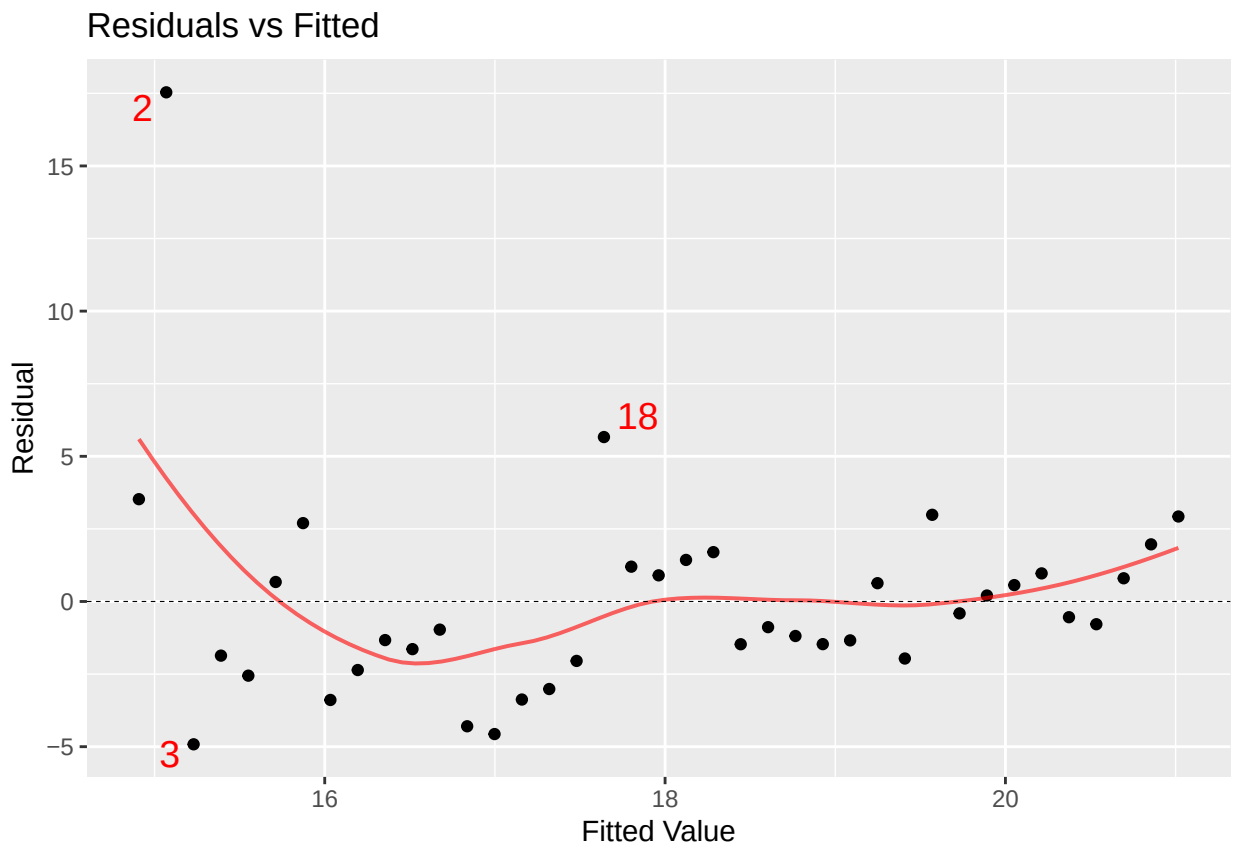
```
gm1 <- lm(anxiety_percentage ~ WEEK, data = mental_health)

msummary(gm1)
```

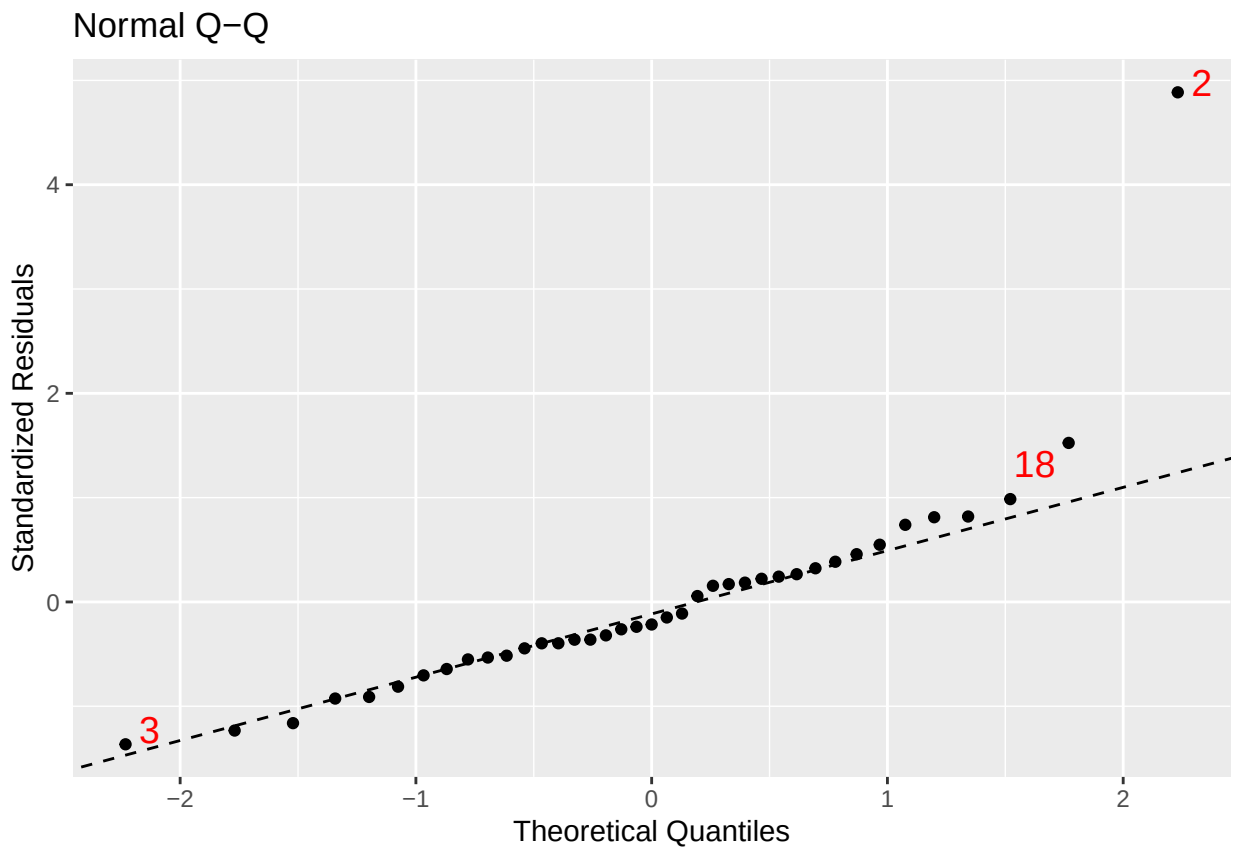
```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.74728    1.22928  11.997 2.57e-14 ***
## WEEK         0.16075    0.05357   3.001   0.0048 **
##
## Residual standard error: 3.765 on 37 degrees of freedom
## Multiple R-squared:  0.1958, Adjusted R-squared:  0.174
## F-statistic: 9.006 on 1 and 37 DF,  p-value: 0.004797
```

```
mplot(gm1, which = 1)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
mplot(gm1, which = 2)
```



> How has mental looked during Covid-19? * tThe general trend of percentage of people who experienced symptoms of anxiety during the 39 weeks of covid is generally positive * The average percent of people who experienced symptoms of anxiety during Covid is 18% * The correlation coefficient (r) between anxiety percentage and week on the national level is 0.44 * There is a weak, positive association between week and anxiety iwith a coefficient of determination (R^2) of 19.6% * This linear model has a residual standard error of 3.765 * Residual standard error is the standard deviation of the residuals, so smaller residual standard errors means predictions are better * In this case, the number is high so this means the predictions are not good * There seems to be a huge deviation from the line of best fit in the early weeks of Covid (2 & 3) * An easy way to detect unequal error variances, outliers, and non-linearity is through a residuals vs. fitted plot * As you can see in the residuals vs. fitted plot, weeks 2, 3, and 18 stand out while the rest of the points are randomly scattered above and below the 0 line * This suggests that the relationship may not be linear, which is also shown in the normal Q-Q plot where the line deviates from the fitted line toward the right

> What does this mean? * This shows that mental health varied widely and was worse in the early weeks of Covid * Towards the later weeks, there did not seem to be any outliers in terms of variance * On average, mental health increased slightly