

Análise Estatística do casos de cancer antes e depois do acidente de Chernobyl

Huggo Parcelly - Victor Paiva

2022-11-21

Descrição do problema

O acidente com a usina nuclear de Chernobyl aconteceu no dia 26 de abril de 1986, no qual houve uma explosão em um reator, lançando material radioativo na atmosfera. A radiação logo se espalhou pelo o mundo, contudo as consequências foram mais profundas sobretudo para três países: Ucrânia, Bielorrússia e Rússia.

Uma consequência grave do acidente foi o aumento dos casos de cancer, que será o objeto de estudo desse relatório. Para tal, utilizou-se um banco de dados que foi retirado a partir de um estudo com as populações das cidades de Gomel, Mogilev e Belarus, que são próximas ao acidente.

Para fins de estudos desse relatório, levantou-se a problemática para comparar as médias dos casos de cancer nos anos de 1985 e 1986, ou seja, antes e depois do acidente.

População associada

Homens e mulheres das cidades de Gomel, Mogilev e Belarus

Variáveis e tipos de interesse para pesquisa

- A variável em estudo é a quantidade de casos de cancer, na população das três cidades e é do tipo quantitativa discreta.

Objetivos do estudo

Objetivos gerais

Comparar as médias dos casos de cancer na população das três cidades, analisando os anos 1985 e 1986.

Objetivos específicos

Realizou-se uma comparação das médias dos casos de cancer, por 100.000 pessoas, das cidades de Gomel, Mogilev e Belarus, nos anos de 1985 e 1986. O objetivo é verificar que houve um aumento na média dos casos e em quanto foi o aumento, através dos testes t student e do Man-whitney.

Hipóteses

Coleta de amostra

Foram realizadas 84 observações, sendo 42 no ano de 1985 e 42 no ano de 1986, período antes e depois do acidente de Chernobyl. Para esses dados foram observados as quantidades de casos, para 7 tipos de câncer, por 100.000 habitantes, homens e mulheres, em mesma proporção.

Os tipos de câncer observados foram colon, rim, pele, pulmão, pancreas, reto e tireoide.

Parâmetro, Estatística e Estimativa

- Parâmetros: Médias dos casos antes e depois da explosão.
- Estatística:
- Estimativa: Estima-se que as médias dos casos de cancer aumentou.

Análise Exploratória de Dados (EDA)

```
if(!require(magrittr)) install.packages("magrittr")
```

```
## Loading required package: magrittr
```

```
library(magrittr)
```

```
if(!require(dplyr)) install.packages("dplyr")
```

```
## Loading required package: dplyr
```

```
##
```

```
## 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
```

```
library(dplyr)
```

```
if(!require(ggplot2)) install.packages("ggplot2")
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
```

```
if(!require(readr)) install.packages("readr")
```

```
## Loading required package: readr
```

```
library(readr)
```

```
dados <- read_csv("cancer_before_and_after_Chernobyl_Belarus.csv")
```

```
## Rows: 84 Columns: 5
```

```
## -- Column specification -----  
## Delimiter: ","  
## chr (3): sex, location, cancer_type  
## dbl (2): year, cases  
##  
## 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.
```

```
## Dados antes do acidente
```

```
before = subset(dados, dados$year == 1985)
```

```
## Dados depois do acidente
```

```
after = subset(dados, dados$year == 1986)
```

```
total_rate_before = sum(before$cases)
```

```
total_rate_after = sum(after$cases)
```

```
total_rate <- c(total_rate_before, total_rate_after)
```

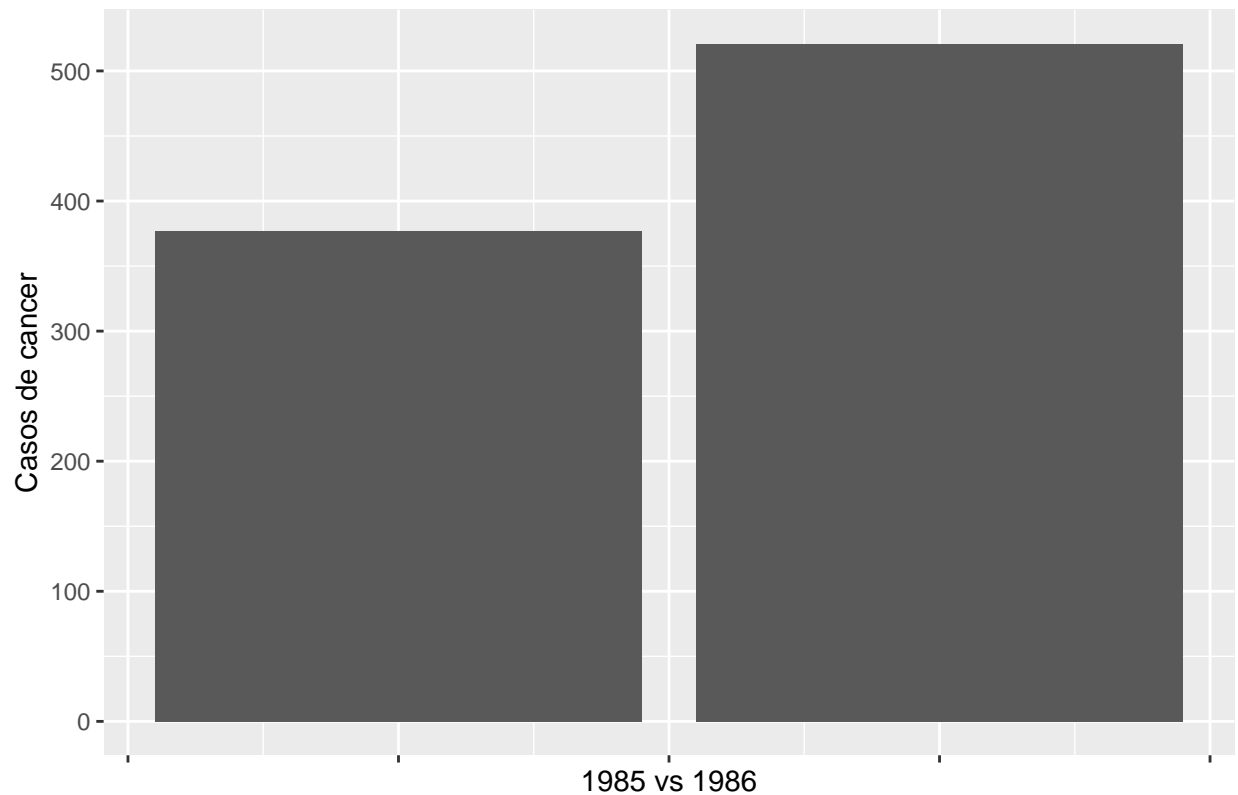
```
years <- c(1985,1986)
```

```
df_total <- data.frame(years, total_rate)
```

Grafico de barras para visualizar a relação de casos de cancer com o acidente nuclear de Chernobyl.

```
df_total %>% ggplot() + geom_col(aes(x=years, y=total_rate)) + ggtitle("Casos de Cancer em Chernobyl")
```

Casos de Cancer em Chernobyl comparando antes e depois do acidente n



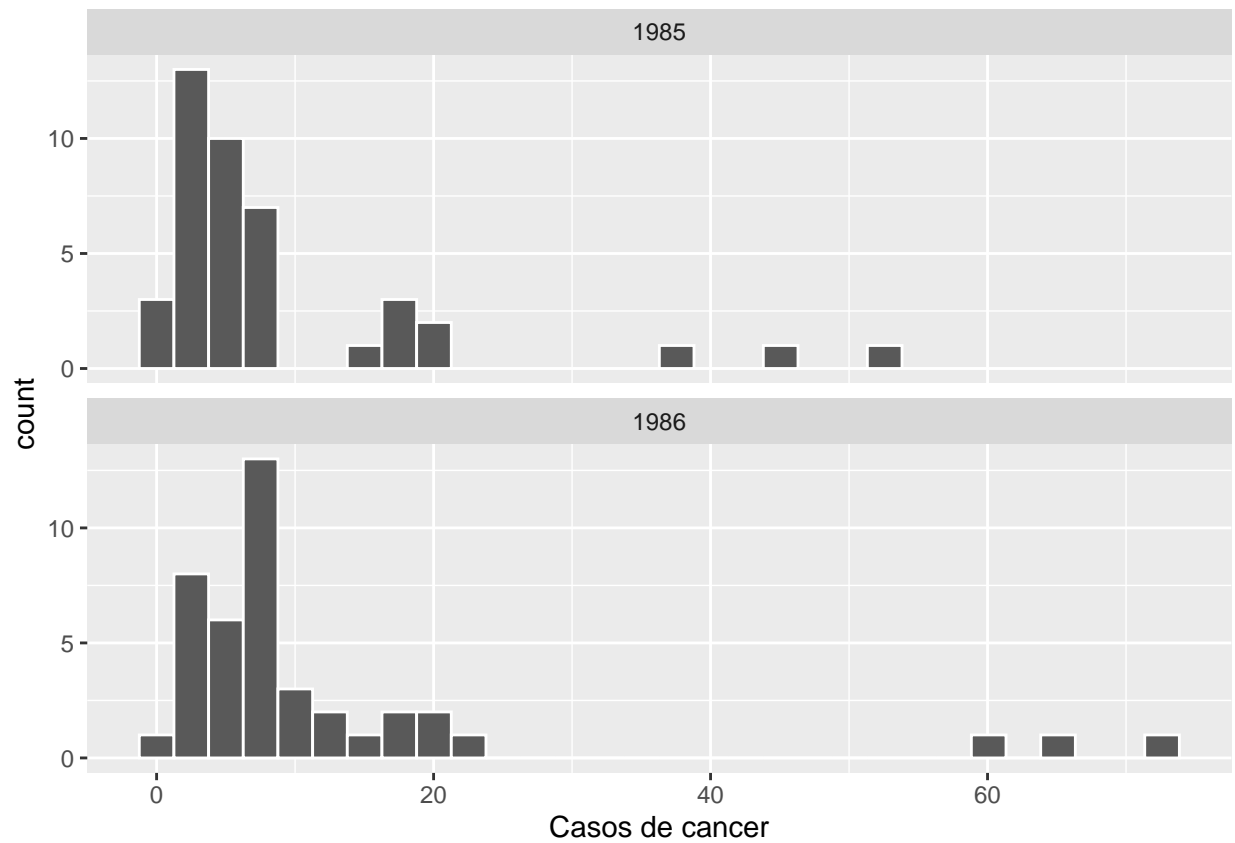
```
perc_higher <- (total_rate_after - total_rate_before)/total_rate_before * 100
```

Houve um aumento de 38,27%(perc_higher) nos casos de câncer.

Histograma separado por ano.

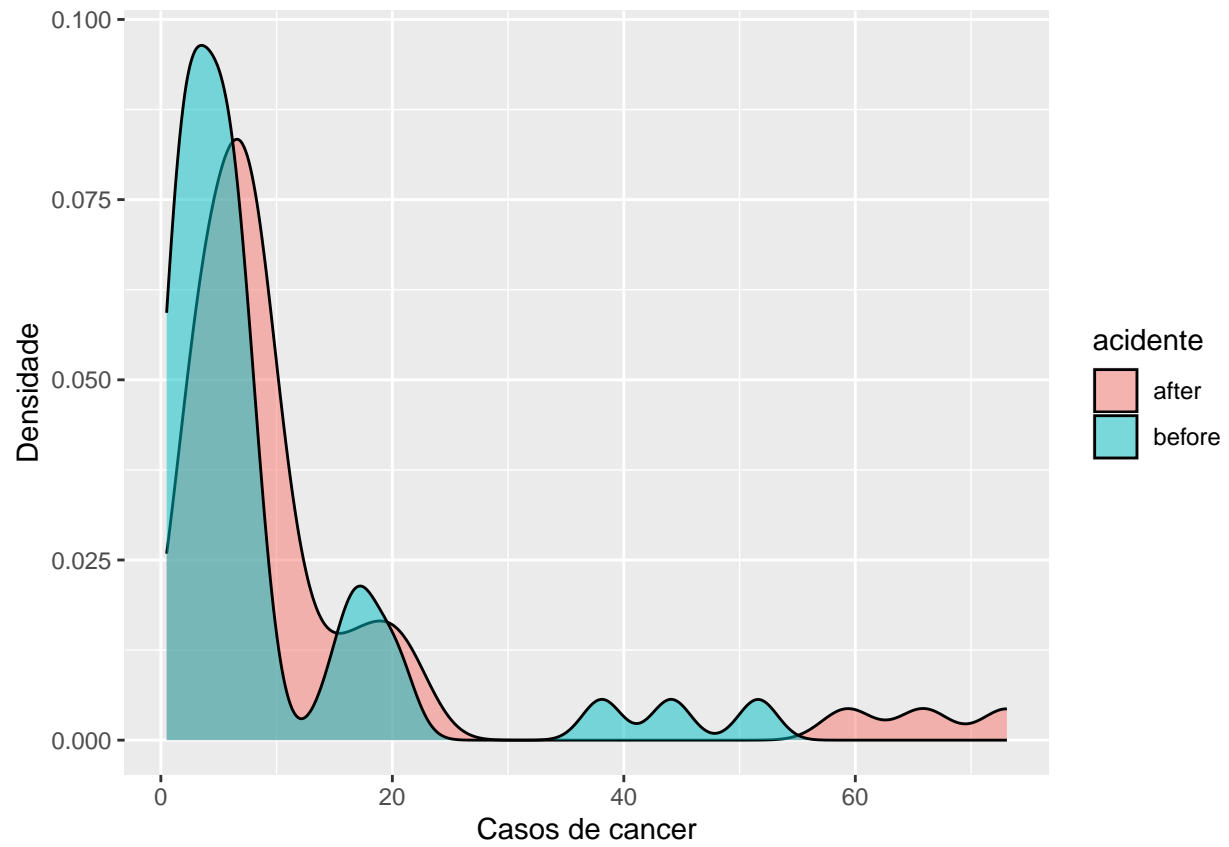
```
ggplot(data = dados, mapping = aes(x = cases)) +  
geom_histogram(color = "white") + facet_wrap(~ year, ncol = 1) +  
labs(x="Casos de cancer")
```

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



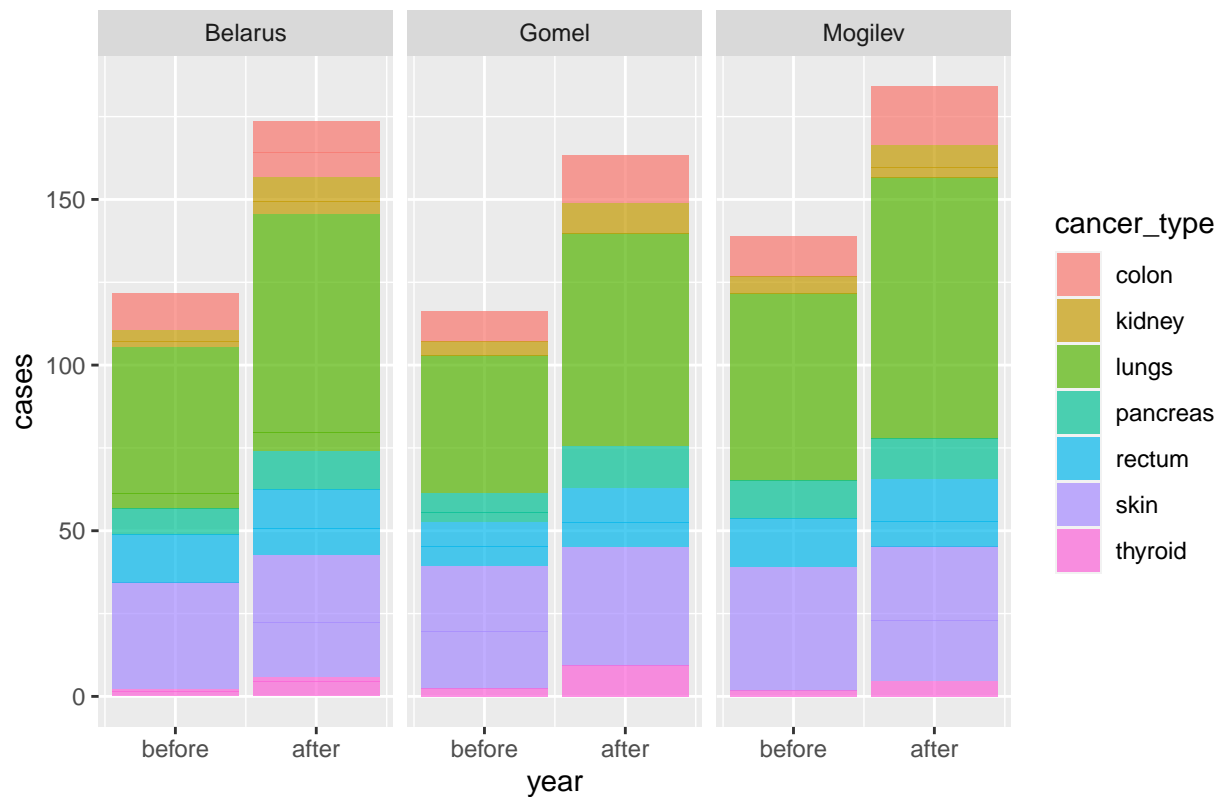
```
dados1 <- transform( dados, acidente= ifelse(year==1985, "before", "after"))

dados1 %>% ggplot(aes(x = cases)) + geom_density(aes(fill = acidente),
alpha = 0.5 ) + labs(x="Casos de cancer", y="Densidade")
```



```
## Gráfico dos casos de cancer nas três cidades, antes e depois do acidente
dados %>% ggplot() + geom_col(aes(year,cases, fill=cancer_type), alpha=0.7) + facet_grid(.~location) + s
```

Casos de cancer em diferentes cidades nantes e depois do acidente nucle



df_total

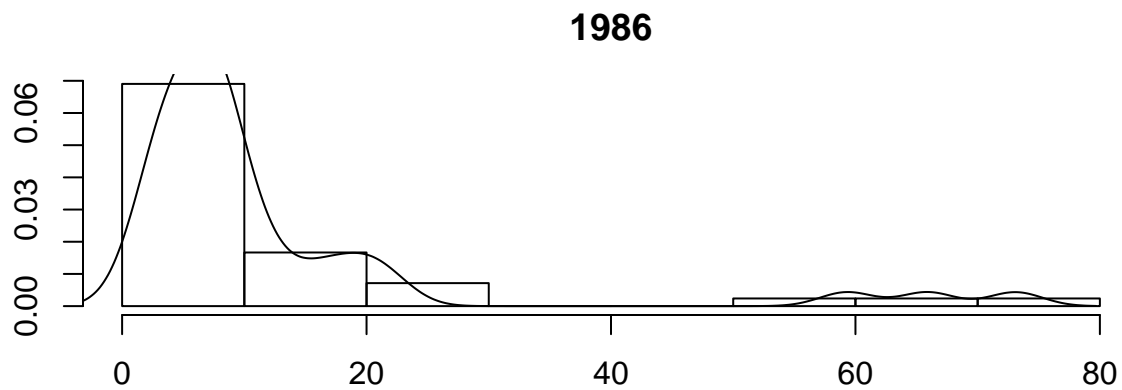
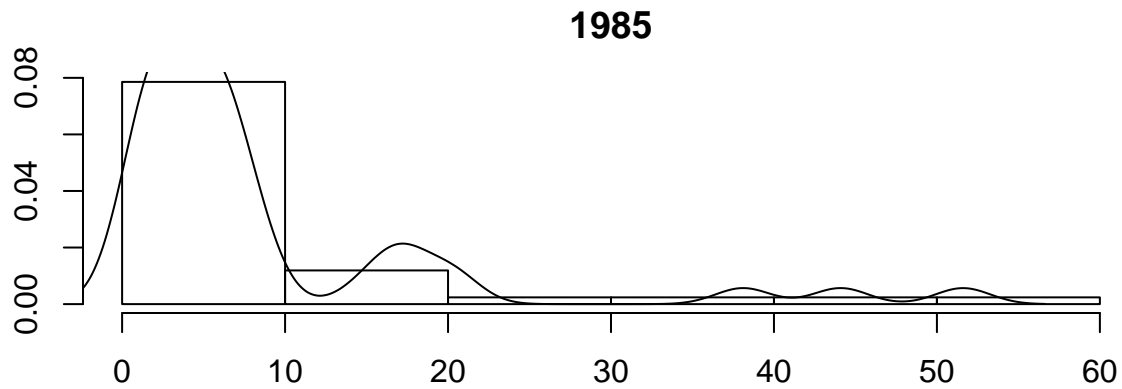
```
##   years total_rate
## 1  1985       376.8
## 2  1986       521.0
```

```
x1 <- dados1$cases[dados1$acidente == "before"]
x2 <- dados1$cases[dados1$acidente == "after"]

par(mfrow = c(2,1), mar = rep(2.5, 4))

hist(x1, prob=TRUE, main = "1985")
lines(density(x1))

hist(x2, prob=TRUE, main = "1986", xlab = "Casos de cancer")
lines(density(x2))
```



Inferências estatísticas

Verificamos a média de todos os casos de câncer nos dois anos, referentes a antes e depois do acidente.

Para comparar as médias utilizamos o test t student para amostras independentes

```
antes <- before$cases
depois <- after$cases
```

```
#test t
```

```
t.test(antes, depois)
```

```
##
##  Welch Two Sample t-test
##
## data:  antes and depois
## t = -1.1319, df = 74.102, p-value = 0.2613
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.477257  2.610590
## sample estimates:
## mean of x mean of y
##  8.971429 12.404762
```


Para comparar as medianas, utilizamos o test de Mann-Whitney

```
test_mann_whitney <- wilcox.test(cases ~ year, data=dados)
```

```
## Warning in wilcox.test.default(x = c(17.3, 16.8, 14.9, 19.9, 20.5, 17.2, :  
## cannot compute exact p-value with ties
```

```
test_mann_whitney
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: cases by year  
## W = 633.5, p-value = 0.02649  
## alternative hypothesis: true location shift is not equal to 0
```

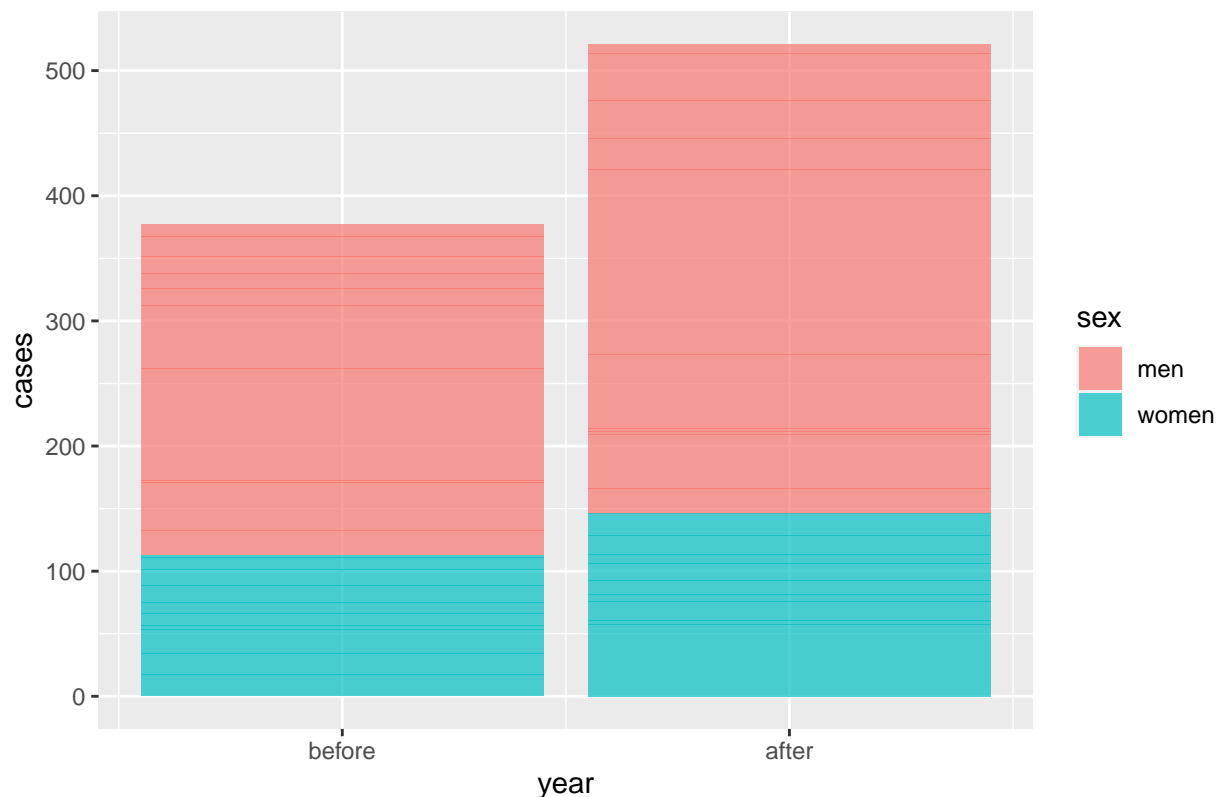
Tomando

H0: mediana dos casos em 1985 = mediana dos casos em 1986 -> $p > 0.05$ H1: mediana dos casos em 1985 != mediana dos casos em 1986 -> $p \leq 0.05$

Portanto, como o valor de $p < 0.05$, então devemos considerar a hipótese alternativa e rejeitar a hipótese nula.

```
## Casos de cancer em homens e mulheres, antes e depois do acidente  
dados %>% ggplot() + geom_col(aes(year, cases, fill=sex), alpha=0.7) + scale_x_continuous(breaks = c(1985, 1986))
```

Casos de cancer em homens e mulheres antes e depois do acidente



```

## Dados de homens antes e depois do acidente
before = subset(dados, dados$year == 1985)
after = subset(dados, dados$year == 1986)

## Homens antes
before_men = subset(before, sex == 'men')
total_rate_before_men = sum(before_men$cases)

## Homens depois
after_men = subset(after, sex == 'men')
total_rate_after_men = sum(after_men$cases)

## Mulheres antes
before_women = subset(before, sex == 'women')
total_rate_before_women = sum(before_women$cases)

## Mulheres depois
after_women = subset(after, sex == 'women')
total_rate_after_women = sum(after_women$cases)

## Porcentagem de aumento de casos em homens e mulheres
total_rate_men <- c(total_rate_before_men, total_rate_after_men)
total_rate_women <- c(total_rate_before_women, total_rate_after_women)

perc_higher_men <- (total_rate_after_men - total_rate_before_men)/total_rate_before_men * 100
perc_higher_men

## [1] 41.91594

perc_higher_women <- (total_rate_after_women - total_rate_before_women)/total_rate_before_women * 100
perc_higher_women

## [1] 29.72493

```

O aumento percentual do número de casos em homens é de 41.92% e em mulheres é de 29.72%

De acordo com os dados, o aumento percentual no número de casos em homens se mostrou maior que em mulheres, indicando que os homens foram mais propensos a contrair câncer que as mulheres.

O que podemos concluir é que após o acidente na usina nuclear de Chernobyl, houve um aumento nos casos de câncer nas cidades próximas. Os dados mostraram uma maior ocorrência de câncer em homens.