

Covid_R_script2

2025-05-29

To start we read in the data and put the tidyverse package in the library. Remember you must install this package once and only once on your machine.

```
Covid_data <- read.csv("covid_deaths_by_age_and_vaccination.csv")
View(Covid_data)

unique(Covid_data)
```

```
# install.packages("tidyverse")
library(tidyverse)
```

Now we create two new data frames filtering by death and vaccination status. We can look in the Environment window in R Studio to see the number of observations for these two (253 and 481). Then two more data frames filtering by just vaccination status.

```
vax_d <- filter(Covid_data, outcome == "death", vaccine_status == "vaccinated")
length(vax_d$outcome)

nvax_d <- filter(Covid_data, outcome == "death", vaccine_status == "unvaccinated")
length(nvax_d$outcome)

vax <- filter(Covid_data, vaccine_status == "vaccinated")
length(vax$outcome)

nvax <- filter(Covid_data, vaccine_status == "unvaccinated")
length(nvax$outcome)
```

Now we compute the death rates in a new data frame, and make the first basic barplot.

```
df <- data.frame(unvaccinated = 253/length(nvax$outcome),
                 vaccinated = 481/length(vax$outcome))
df

# actually make it a data.frame with one column
dfv <- data.frame(death_rates = c(round(253/length(nvax$outcome)*100, 2),
                                round(481/length(vax$outcome)*100, 2)))
rownames(dfv) = c("unvaccinated", "vaccinated")
dfv
barplot(dfv$death_rates, col = "deepskyblue", ylab = "Percentages",
        names.arg = rownames(dfv),
        main = "Death Rate by Vaccination Status")
```

To get the data labels we need the x-coordinates of the bars. The barplot as a data object is actually a vector with these x-coordinates!

```

p <- barplot(dfv$death_rates, col = "deepskyblue", ylab = "Percentages",
             names.arg = rownames(dfv),
             main = "Death Rate by Vaccination Status",
             ylim = c(0, max(dfv$death_rates)+0.1))
data <- dfv$death_rates

# p is a vector with the x-coords of the bars
p
data

barplot(dfv$death_rates, col = "deepskyblue", ylab = "Percentages",
        names.arg = rownames(dfv),
        main = "Death Rate by Vaccination Status",
        ylim = c(0, max(dfv$death_rates)+0.1))
text(x = p, y = data - 0.025, labels = as.character(data))

```

We will recreate the figure with ggplot using tables.

```

table(Covid_data$outcome, Covid_data$vaccine_status)

# let's play with tables
tab <- table(Covid_data$outcome, Covid_data$vaccine_status)
class(tab)

tab[1,1]
tab[,1]

sum(tab[,1])

```

We compute the survival rates using a for loop.

```

ncol(tab)
perc_tab <- matrix(0, ncol = ncol(tab), nrow = nrow(tab))
perc_tab
for (j in 1:ncol(tab)) {
  perc_tab[,j] <- tab[,j]/sum(tab[,j])
  perc_tab[,j] <- round(perc_tab[,j]*100, 2)
}
perc_tab

```

Now we turn these rates again into a data frame and create the first target graphic.

```

dfperc <- data.frame("death.rate" = perc_tab[1, ], "survival.rate" = perc_tab[2, ])
rownames(dfperc) <- c("unvaccinated", "vaccinated")
dfperc

# Finally we get target graphic #1
ggplot(data = dfperc, mapping = aes(x = rownames(dfperc), y = death.rate)) +
  geom_bar(stat = "identity", fill = "blue") +
  ylim(0, max(dfperc$death.rate)+0.1) +
  xlab(NULL) + ylab("Percentage") + ggtitle("Death Rate by Vaccination Status") +
  geom_text(aes(label = death.rate), vjust = 1.5, color = "white")

```

To create the last two graphics we can use the prop.table function.

```
table(Covid_data)
tbl <- table(Covid_data)
dimnames(tbl)

# Now use proptable to get percentages
prop.table(tbl, margin = c(1,2))
round(prop.table(tbl, margin = c(1,2))*100, 2)

ptbl <- round(prop.table(tbl, margin = c(1,2))*100, 2)
ptbl[1:4]
```

We can make the second target graphic using these values from the prop.table and creating a new data frame that has tidy data.

```
# Let's make the second target graphic
pdf2 <- data.frame(death_rates = c(ptbl[1],ptbl[3],ptbl[2],ptbl[4]),
                  Age_group = c("50 +", "50 +","under 50","under 50"),
                  Vax_status = c("unvaccinated", "vaccinated",
                                "unvaccinated","vaccinated"))

pdf2

ggplot(pdf2, aes(x = Age_group, y = death_rates, fill = Vax_status)) +
  geom_bar(stat = "identity", position = "dodge") +
  ylim(0, max(pdf2$death_rates)+0.1) +
  ylab("Death Rates %") +
  scale_fill_manual(values = c("blue", "red")) +
  geom_text(aes(label = death_rates),position = position_dodge(0.9),
            vjust = -0.5, size = 3, color = "black")
```

And now we can create the third target graphic.

```
# Compute relative risk
# first create pdf for props and graph!
ptbl0 <- prop.table(tbl, margin = c(1,2))
pdf <- data.frame(rel_risk = round(c(ptbl0[1]/ptbl0[3],
                                   ptbl0[2]/ptbl0[4]), 2))

row.names(pdf) = c("50 +", "under 50")
pdf

ggplot(data = pdf, mapping = aes(x = rownames(pdf), y = rel_risk)) +
  geom_bar(stat = "identity", fill = "blue") +
  ylim(0, max(pdf$rel_risk)+0.1) +
  xlab(NULL) + ylab(NULL) + ggtitle("Relative Risk for Unvaccinated") +
  geom_text(aes(label = rel_risk), vjust = 1.5, color = "white") +
  geom_hline(yintercept = 1, linetype = "dashed", size = 2) +
  annotate("text", x=0.9, y = 1.1, label = "Vaccine has no effect",
          color = "white") +
  annotate("rect", xmin=1, xmax=2, ymin=0.69, ymax=0.9, fill="white") +
  annotate("text", x=1.5, y = 0.8, label = "Unvaccinated less likely to die") +
  annotate("rect", xmin=1, xmax=2, ymin=1.5, ymax=1.71, fill="white") +
```

```
annotate("text", x=1.5, y = 1.61, label = "Unvaccinated more likely to die") +  
annotate("segment", x=1.5, xend=1.5, y=0.69, yend=0.3, arrow=arrow(),size = 2) +  
annotate("segment", x=1.5, xend=1.5, y=1.71, yend=2.1, arrow=arrow(),size = 2)
```

The End.