

Practice 2

Ekta Chaudhary

17/07/2020

```
china_data = read_sas("./data/corona_china.sas7bdat") %>%
  janitor::clean_names()
southkorea_data = read_sas("./data/corona_southkorea.sas7bdat") %>%
  janitor::clean_names()
```

Data set:

Variable Name

ID Patient unique ID Agegroup Age (years) at diagnosis. 0 = 0-9 years 1 = 10-19 years 2 = 20-29 years 3 = 30-39 years 4 = 40-49 years 5 = 50-59 years 6 = 60-69 years 7 = 70-79 years 8 = 80+ years Sex 1 = female 2 = male Hypertension Comorbid hypertension at COVID19 diagnosis 1 = yes 0 = no CVD Comorbid cardiovascular disease at COVID19 diagnosis 1 = yes 0 = no Dead 1 = yes 0 = no

Number of people in each gender

```
china_data %>%
  group_by(
    sex
  ) %>%
  dplyr::summarise(
    number = n()
  ) %>%
  knitr::kable()
```

| sex | number |
|-----|--------|
| 1 | 21691 |
| 2 | 22981 |

Number of people in each age group

```
china_data %>%
  group_by(
    agegroup
  ) %>%
  dplyr::summarise(
    number = n()
  ) %>%
  knitr::kable()
```

| agegroup | number |
|----------|--------|
| 0 | 416 |
| 1 | 549 |
| 2 | 3619 |
| 3 | 7600 |

| agegroup | number |
|----------|--------|
| 4 | 8571 |
| 5 | 10008 |
| 6 | 8583 |
| 7 | 3918 |
| 8 | 1408 |

Number of people with and without hypertension

```
china_data %>%
group_by(
  hypertension
) %>%
dplyr::summarise(
  number = n()
) %>%
dplyr::mutate(
  percent = (number/sum(number))*100
) %>%
knitr::kable()
```

| hypertension | number | percent |
|--------------|--------|----------|
| 0 | 38954 | 87.20004 |
| 1 | 5718 | 12.79996 |

Percentage of people with and without CVD

```
china_data %>%
group_by(
  cvd
) %>%
dplyr::summarise(
  number = n()
) %>%
dplyr::mutate(
  percent = (number/sum(number))*100
) %>%
knitr::kable()
```

| cvd | number | percent |
|-----|--------|-----------|
| 0 | 42796 | 95.800501 |
| 1 | 1876 | 4.199499 |

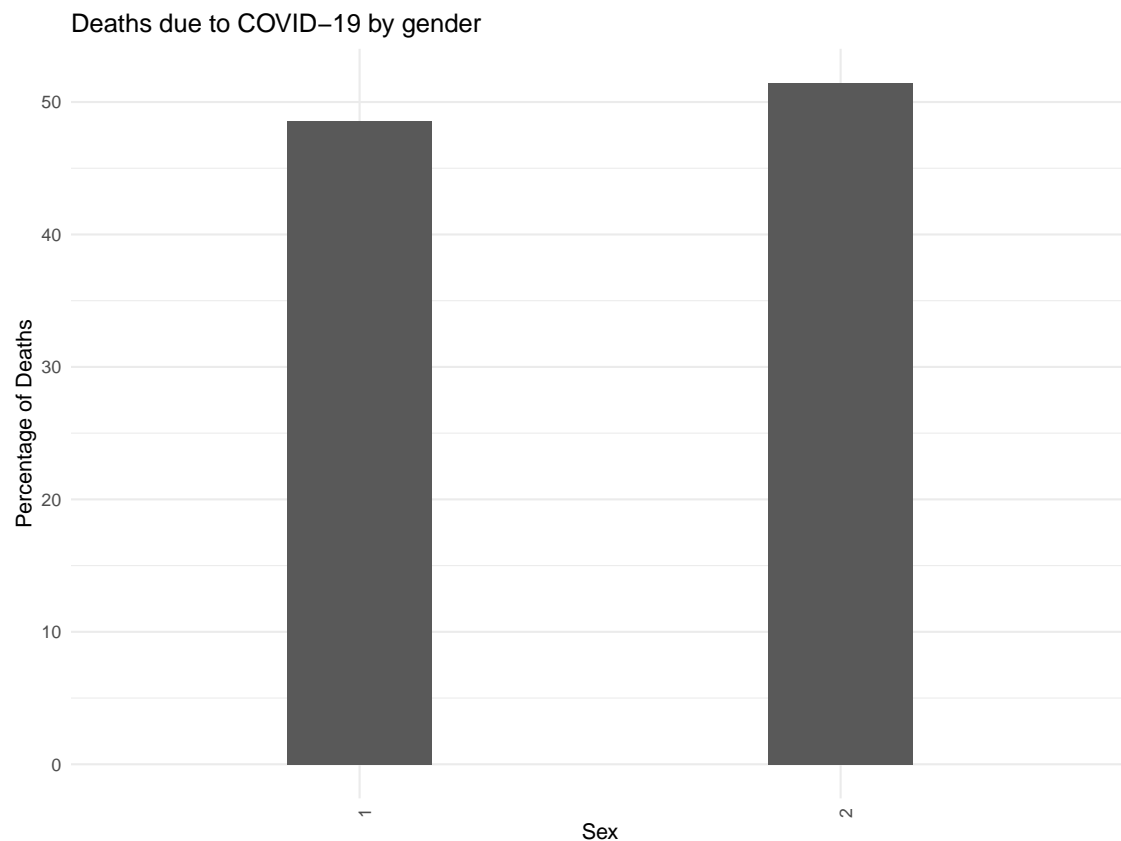
Deaths due to COVID-19 by gender

```
china_data %>%
group_by(
  sex
```

```

) %>%
dplyr::summarise(
  count = n()
) %>%
dplyr::mutate(
  percent = (count/sum(count))*100
) %>%
ggplot(
  aes(
    x = factor(sex), y = percent
  )
) + geom_bar(stat = "identity", width = 0.3) + labs(x = "Sex", y = "Percentage of Deaths", title = "Deaths due to COVID-19 by gender")

```



Percentage of COVID-19 deaths by gender in people with and without Hypertension

```

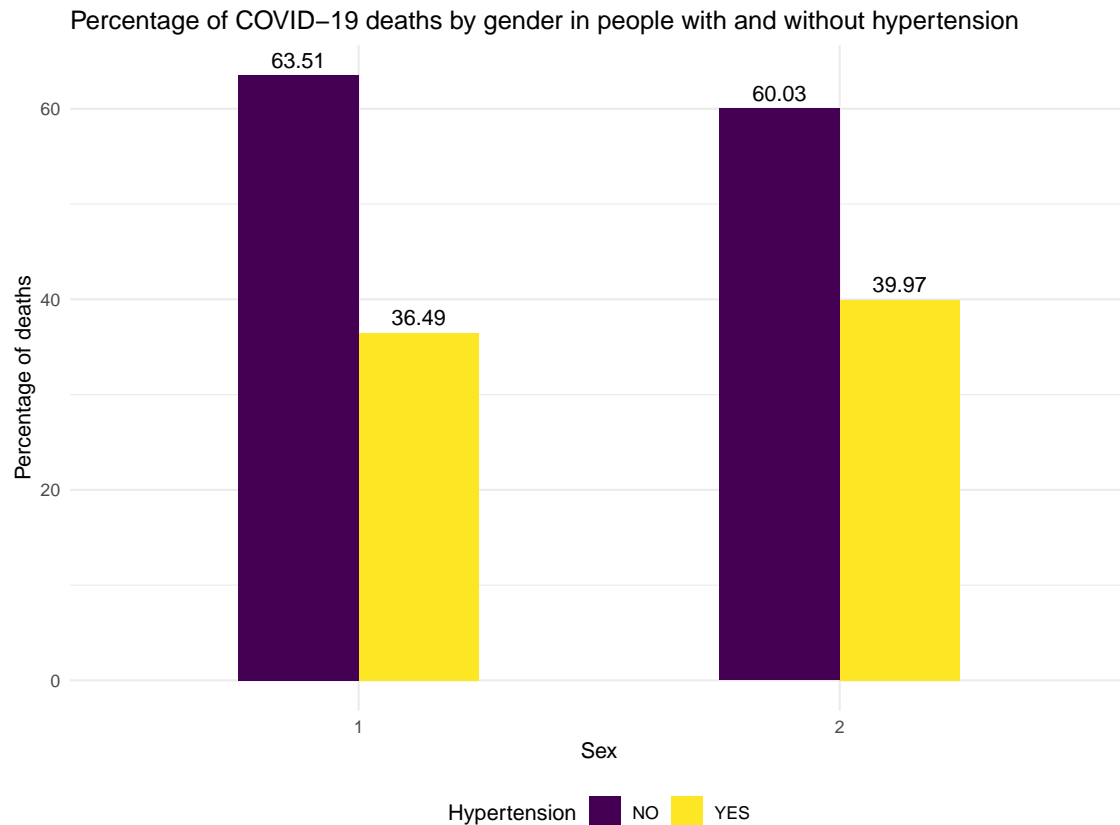
china_data %>%
  filter(
    dead == 1
  ) %>%
  mutate(
    sex = factor(sex),
    hypertension = factor(hypertension)
  ) %>%
  group_by(
    sex, hypertension
  ) %>%
  dplyr::summarise(

```

```

count = n()
) %>%
dplyr::mutate(
  percent = round((count/sum(count))*100, digits = 2)
) %>%
ggplot(aes(x = sex, y = percent)) +
geom_bar(aes(fill = hypertension), position = "dodge", stat = "identity", width = .5) +
  geom_text(aes(label = percent, group = hypertension), position = position_dodge(width = 0.5), vjust = -1)
labs(x = "Sex", y = "Percentage of deaths", title = "Percentage of COVID-19 deaths by gender in people with and without hypertension")

```



Percentage of COVID-19 deaths by gender in people with and without CVD

```

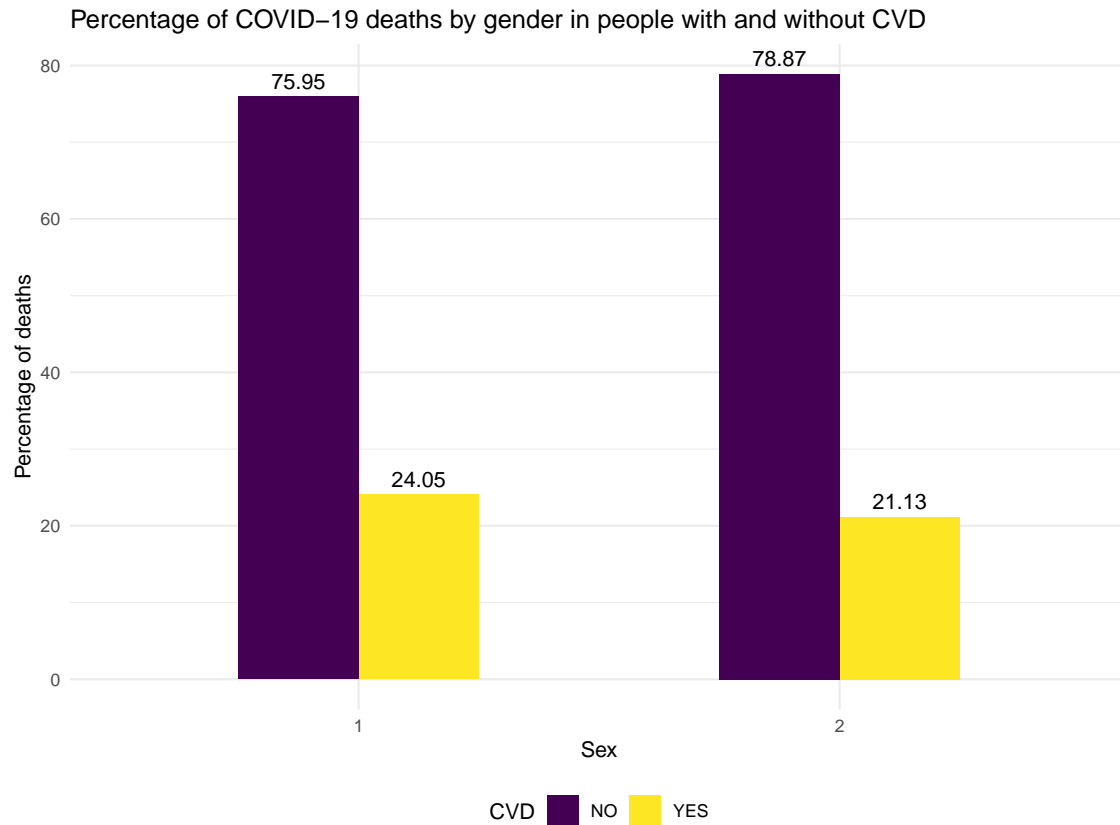
china_data %>%
  filter(
    dead == 1
  ) %>%
  mutate(
    sex = factor(sex),
    cvd = factor(cvd)
  ) %>%
  group_by(
    sex, cvd
  ) %>%
  dplyr::summarise(
    count = n()
  ) %>%
  dplyr::mutate(

```

```

percent = round((count/sum(count))*100, digits = 2)
) %>%
ggplot(aes(x = sex, y = percent)) +
geom_bar(aes(fill = cvd), position = "dodge", stat = "identity", width = .5) +
geom_text(aes(label = percent, group = cvd), position = position_dodge(width = 0.5), vjust = -0.5)
labs(x = "Sex", y = "Percentage of deaths", title = "Percentage of COVID-19 deaths by gender in p

```



Number of COVID-19 deaths by age groups:

```

china_data %>%
  select(
    agegroup, dead
  ) %>%
  group_by(
    agegroup
  ) %>%
  dplyr::summarise(
    count = n()
  ) %>%
  knitr::kable()

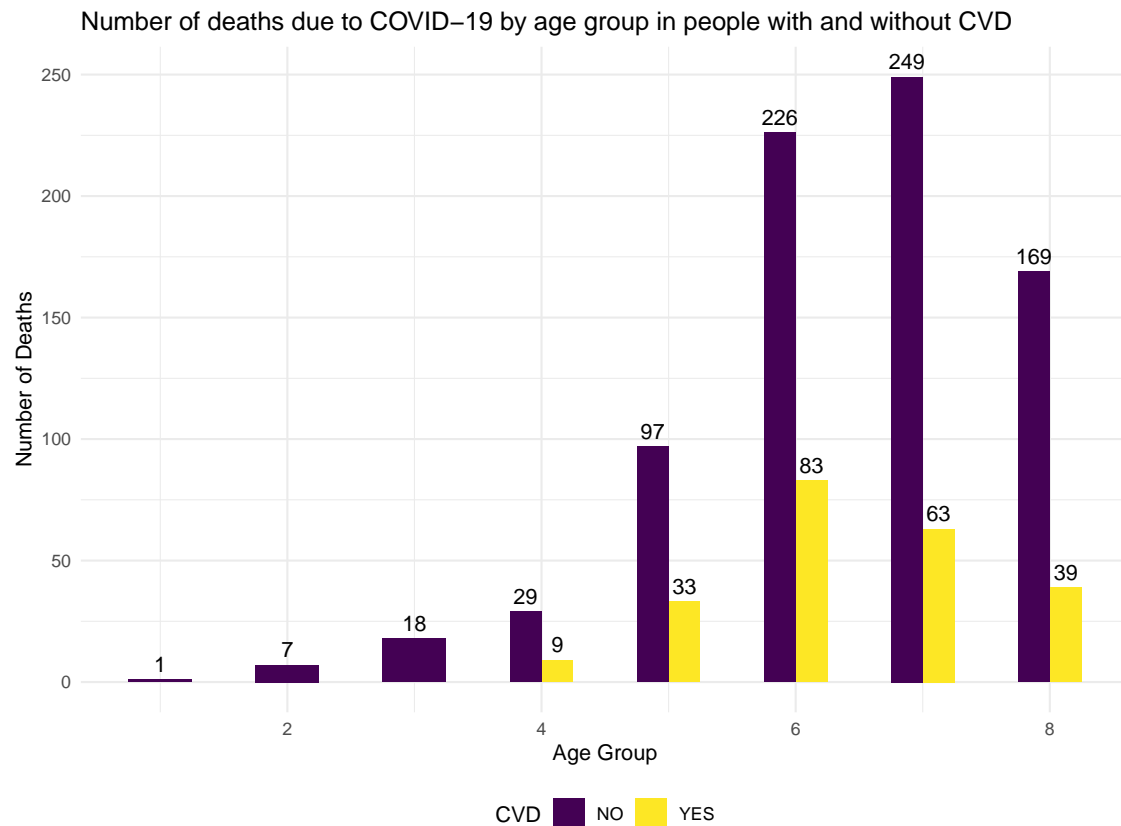
```

| agegroup | count |
|----------|-------|
| 0 | 416 |
| 1 | 549 |
| 2 | 3619 |
| 3 | 7600 |

| agegroup | count |
|----------|-------|
| 4 | 8571 |
| 5 | 10008 |
| 6 | 8583 |
| 7 | 3918 |
| 8 | 1408 |

Number of deaths due to COVID-19 by age group in people with and without CVD

```
china_data %>%
  filter(
    dead == 1
  ) %>%
  mutate(
    cvd = factor(cvd)
  ) %>%
  group_by(
    agegroup, cvd
  ) %>%
  dplyr::summarise(
    count = n()
  ) %>%
  ggplot(
    aes(
      x = agegroup, y = count
    )
  ) + geom_bar(aes(fill = cvd), position = "dodge", stat = "identity", width = .5) +
    geom_text(aes(label = count, group = cvd), position = position_dodge(width = 0.5), vjust = -0.5)
```



Number of deaths due to COVID-19 by age group in people with and without Hypertension

```
china_data %>%
  filter(
    dead == 1
  ) %>%
  mutate(
    hypertension = factor(hypertension)
  ) %>%
  group_by(
    agegroup, hypertension
  ) %>%
  dplyr::summarise(
    count = n()
  ) %>%
  ggplot(
    aes(
      x = agegroup, y = count
    )
  ) +
  geom_bar(aes(fill = hypertension), position = "dodge", stat = "identity", width = .5) +
  geom_text(aes(label = count, group = hypertension), position = position_dodge(width = 0.5), vjust
```

