Measurement 2 - tidyverse

Anna Yorozuya

University of Tokyo

June 7, 2022

Table of Contents

How to save/print graphs

ggsave

- `ggsave(path, filename, extension)`
- for example, if you want to save the figure as a pdf in the result_figures directory, `ggsave("results_figures/education_by_province.pdf")`

gridExtra

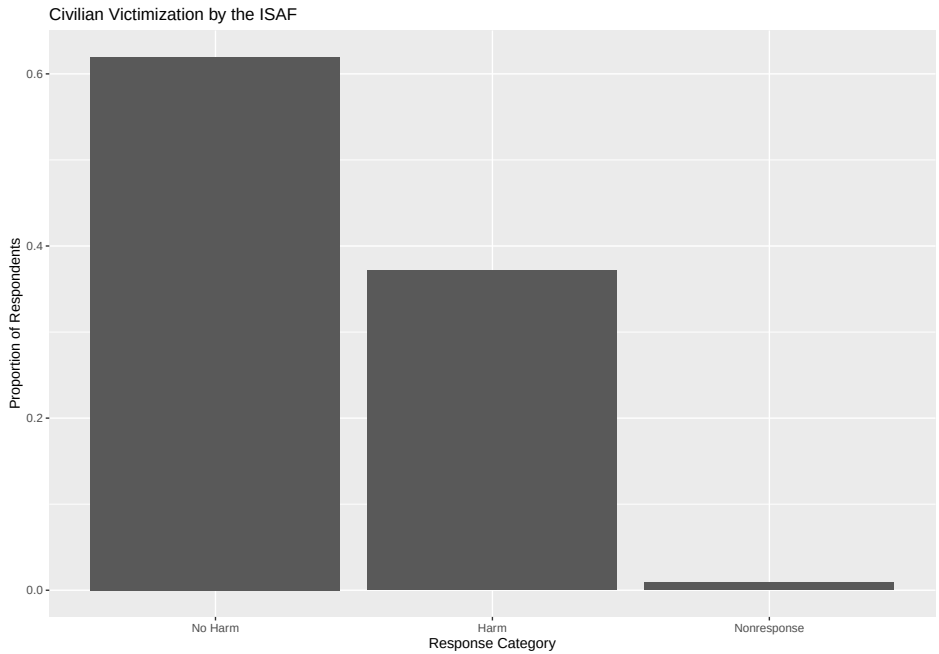
- save multiple plots into a single file
- first, load the package with `library(gridExtra)`
- use the `grid_arrange()`

Section 1

Review: ggplot2

Example: Bar plot (basic)

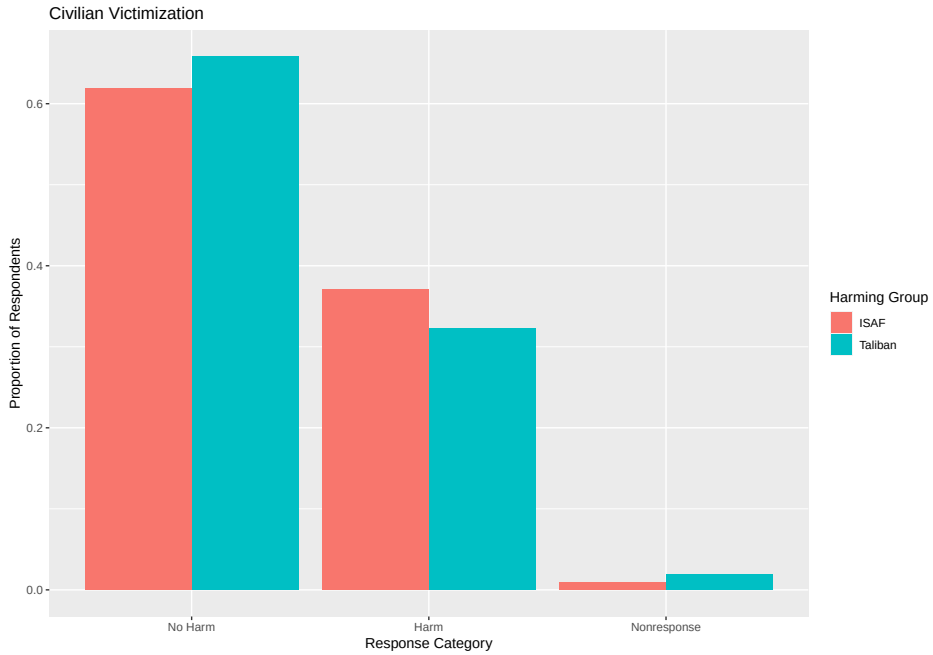
```
ggplot(data = afghan, # Tell R what data to use
    aes(x = as.factor(violent.exp.ISAF))) + # specify the x-axis
  geom_bar(aes(y = stat(prop), # add a bar plot layer
           group = 1)) +
  scale_x_discrete(labels = c("No Harm","Harm","Nonresponse")) +
  ylab("Proportion of Respondents") + # Add a label to y-axis
  xlab("Response Category") + # Add a label to the x-axis
  ggtitle("Civilian Victimization by the ISAF") # Add a title
```



- `aes(y = stat(prop))`: the y-axis shows the proportion, not the count. this is equivalent to `aes(y = ..prop..)`
- `aes(group = 1)`: plot the proportion of the total
- `scale_x_discrete()`: indicate which value to display in the x-axis

Example: Bar plot (advanced)

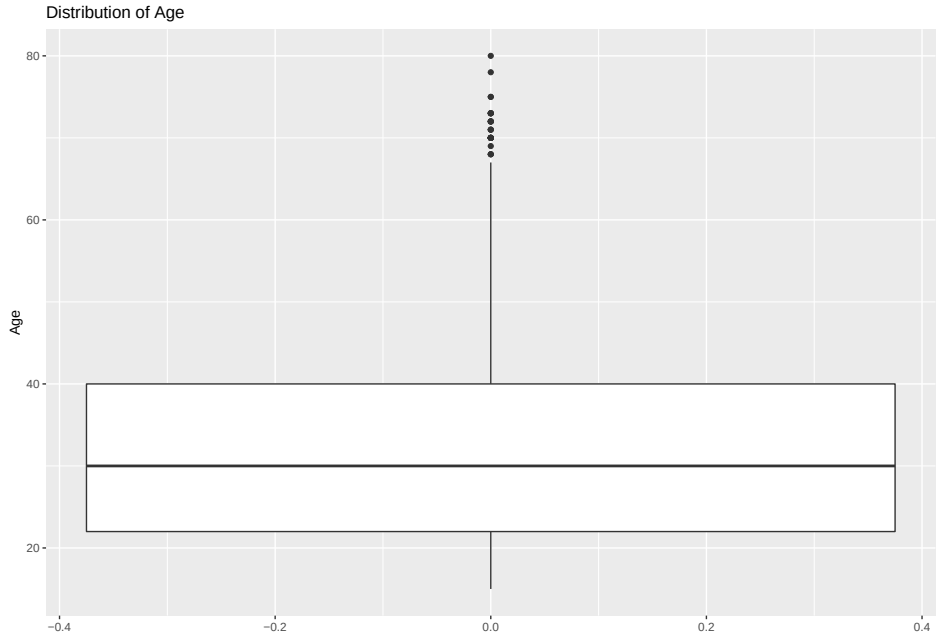
```
ggplot(data = afghan_reshape,
       aes(x = as.factor(harm))) +
geom_bar(aes(y = ..prop.., fill = harming_group,
             group = harming_group),
             position = "dodge") +
  scale_x_discrete(labels = c("No Harm","Harm","Nonresponse")) +
  scale_fill_discrete(name = "Harming Group", labels = c("ISAF", "Taliban")) +
  ylab("Proportion of Respondents") +
  xlab("Response Category") +
  ggtitle("Civilian Victimization")
```



- `position = "dodge"`: avoid overlapping, places bars side by side
- `scale_fill_discrete()`: change the labels on the legend for the bar colors

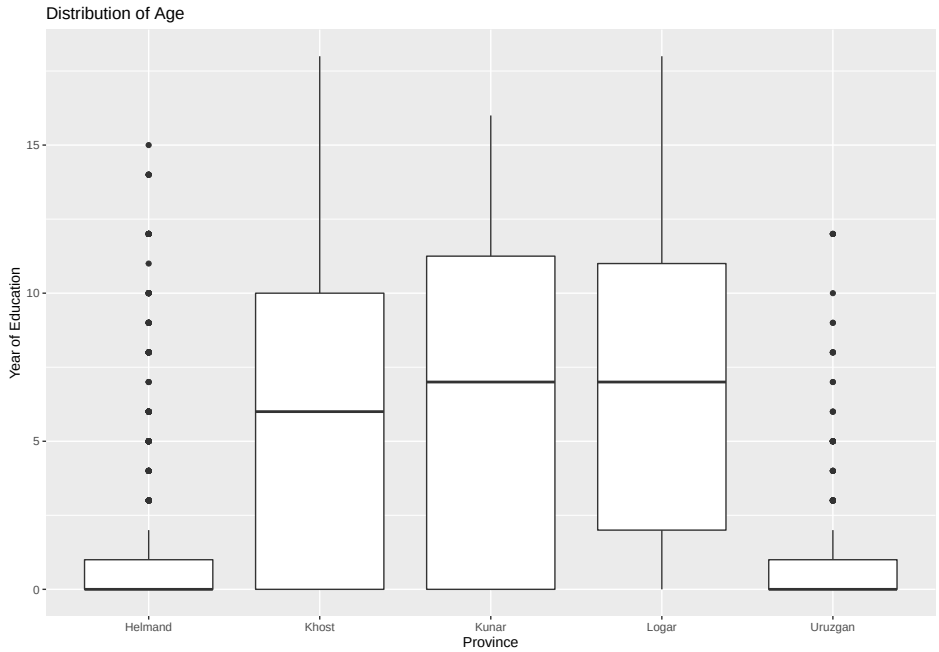
Example: Boxplot (basic)

```
ggplot(afghan, aes(y = age)) +
  geom_boxplot() +
  labs(y = "Age", x = "", title = "Distribution of Age")
```



Example: Boxplot (advanced)

```
ggplot(afghan, aes(y = educ.years, x = province)) +
  geom_boxplot() +
  labs(y = "Year of Education", x = "Province", title = "Distr
```



- `aes(x = province)`: create boxplot for each value of province

tidymodels package

What is tidymodels?

- "a collection of packages for modeling and machine learning using tidyverse principles" (official description of this package)
- the `tidy()` function here is contained in the `broom` package as well

function tidy(): convert the output of a model into a tibble

- takes the output of an R model (such as `kmeans()`) as an argument

References

- “Quantitative Social Science: An Introduction” - Kosuke Imai
- “Quantitative Social Science: An Introduction in Tidyverse” - Kosuke Imai and Nora Webb Williams
- R for data science - H.Wickham and G.Grolemund
- Tidymodels package