

Data Summaries & Visualization

John Karuitha

3/18/2021

Describing and Exploring Categorical Data

- ▶ As noted, categorical or qualitative data shows categories/ buckets/ classes.
- ▶ Examples; are height- Short, Medium, Tall; Age- 0 to 10 years, 11-15 years, 16-20 years, etc.
- ▶ Note that the examples above could be quantitative e.g age of three students- 10, 12, 16 years.
- ▶ You must distinguish qualitative and quantitative variables by looking at the context and presentation.

How can we summarise categorical variables

- ▶ We mostly use frequency tables or contingency tables
- ▶ Lets have an example as follows for the ages of senior high school students

```
## # A tibble: 6 x 2
##   name      age
##   <chr>    <chr>
## 1 Paul     10-15
## 2 Catherine 16-20
## 3 Oloo     16-20
## 4 Etyang   16-20
## 5 Nyaga    10-15
## 6 Mwajuma  21-25
```

Summarising categorical variables

- ▶ Note that this data is categorical- students are presented in age groups and not real age.
- ▶ Let us summarise the age groups in tabular format.

Age	Freq
10-15	2
16-20	3
21-25	1

- ▶ We can also do relative frequencies as percentage of the occurrence of each age group.

##

10-15 16-20 21-25

0.3333333 0.5000000 0.1666667

Contingency tables

- ▶ We typically summarise data using contingency tables.
- ▶ Let us have a look at another example.

```
## # A tibble: 6 x 3
##   name      age  sex
##   <chr>    <chr> <chr>
## 1 Paul      10-15 Male
## 2 Catherine 16-20 Female
## 3 Oloo      16-20 Male
## 4 Etyang    16-20 Male
## 5 Nyaga     10-15 Male
## 6 Mwajuma   21-25 Female

##
## Female    Male
##         2      4
```

- ▶ we can have a two way contingency table as follows.

```
##
```

Exercises

- ▶ You are provided with the following categorical dataset for ages of random people.
- ▶ Also remember how to write a character vector in R; Using quotes.

```
ages <- c("Toddler", "Teen", "Teen", "Adult",  
          "Adult", "Elder", "Adult", "Toddler",  
          "Teen", "Teen")
```

```
ages
```

```
## [1] "Toddler" "Teen"    "Teen"    "Adult"   "Adult"   "  
## [8] "Toddler" "Teen"    "Teen"
```

- ▶ Draw a frequency table
- ▶ Draw a relative frequency table

Setting up a project/working directory in R

Intro to R-Markdown