Replicating Analysis and Descriptive Statisitcs

Abbey Skinner

Student Concerns and Perceived Challenges in Introductory Statistics, how the Frequency shifted during COVID-19, and

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.1 --

v ggplot2 3.4.1     v purrr   0.3.5
v tibble  3.1.7     v dplyr   1.0.9
v tidyr   1.2.0     v stringr 1.4.0
v readr   2.1.2     v forcats 0.5.1

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(gmodels)
```

```
data <- read.csv("/Users/abbeyskinner/Dropbox/My Mac (MacBook Pro)/Desktop/Amherst/ Intern
```

```
## Figuring out what the variables are

count(data, urm)  ##  0 = Non-URM; 1 = URM; -99 = MISSING
count(data, gender) ## 0 = Male; 1 = Female; 2 = Non-Binary; -99 = MISSING
count(data, covid) ## 0 = ??; 1 = ?? (either pre-covid or mid-covid)

## CODEBOOK (for survey questions):
```

```
###### -99 = MISSING; 0 = No Concern; 1 = Concern

count(data, cw_r_1) ## R Coding
count(data, cw_u_1) #Understanding concepts
count(data, cw_w_1) # Workload
count(data, vc_1) # virtual learning
count(data, ss_tm_1) # time management
count(data, ss_r_1) # inaccessabile resources
count(data, ss_m_1) # motivation
count(data, ss_fm_1) # fixed mindset
count(data, ss_se_1) # study environment
count(data, le_pk_1) # lack of proir knowledge
count(data, le_ts_1) # transfer student
count(data, e_m_1) # memorizing
count(data, e_p_1) # performance
count(data, pp_p_1) # instructor
count(data, pp_sc_1) # social comparison
count(data, pp_gw_1) # group work
count(data, c_eo_1) # external obligations
count(data, ps_hwb_1) # physical/mental health
count(data, ps_fb_1) # financial burdens
count(data, ps_pls_1) # personal struggles unrelated to health/finances
count(data, ps_fnb_1) # non-belonging & isolation
count(data, ps_pro_1) # pandemic related concerns
count(data, o_n_1) ## no concerns
```

```
table2_data <- data%>%
  select(urm, gender, covid, cw_r_1, cw_u_1, cw_w_1, vc_1, ss_tm_1, ss_r_1, ss_m_1, ss_fm_
  rename(r_coding = cw_r_1, understanding_concepts = cw_u_1, workload = cw_w_1, virtual_le
  filter(r_coding >= 0) ## remove missing values
```

```
gender_table <- table2_data %>%
  select(-c(urm, covid))%>%
  group_by(gender)%>%
  filter(gender >= 0)

gender_table
```

```
# A tibble: 1,259 x 24
# Groups:   gender [3]
   gender r_coding understanding_conc~ workload virtual_learning time_management
```

```
    <int>    <int>               <int>    <int>            <int>           <int>
 1      0        1                   0        0                0               0
 2      0        0                   0        0                0               0
 3      0        1                   0        0                0               0
 4      0        1                   0        1                0               0
 5      0        0                   0        0                0               0
 6      0        1                   0        1                0               0
 7      0        0                   0        0                0               0
 8      0        0                   0        0                0               0
 9      0        0                   0        0                0               0
10      0        1                   0        0                0               0
# ... with 1,249 more rows, and 18 more variables:
#   inaccessabile_resources <int>, motivation <int>, fixed_mindset <int>,
#   study_enviornment <int>, lack_of_prior_knowledge <int>,
#   transfer_student <int>, memorizing <int>, performance <int>,
#   instructor <int>, social_comparison <int>, group_work <int>,
#   external_obligations <int>, physical_mental_health <int>,
#   financial_burdens <int>, ...
```

```r
CrossTable(gender_table$r_coding, gender_table$gender, prop.c=TRUE)
```

```
   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  1259


                     | gender_table$gender
gender_table$r_coding |         0 |         1 |         2 | Row Total |
----------------------|-----------|-----------|-----------|-----------|
                    0 |       247 |       618 |        15 |       880 |
                      |     1.158 |     0.474 |     0.075 |           |
```

	0.281	0.702	0.017	0.699
	0.748	0.680	0.750	
	0.196	0.491	0.012	
1	83	291	5	379
	2.688	1.102	0.173	
	0.219	0.768	0.013	0.301
	0.252	0.320	0.250	
	0.066	0.231	0.004	
Column Total	330	909	20	1259
	0.262	0.722	0.016	

```
## could do this for each question ...  is there a better way?
## the proportions in table 2 are the 4th row in the 2nd block
```