# R Markdown Tutorial

*Eric Karsten*

*25 February, 2018*

## Contents



Now let's say that we want to have a nice PDF readout of some of our tables and analyses in R. The tool we will use is Rmarkdown. This allows you to generate a PDF document with output from R code with very little effort. There are lots of options for how you run your code chunks so that they turn out nicely.

## Including or not including code chunks

Our first code chunk brings in our dataset, so it's not useful to include that in the markdown (and you don't see it below in the PDF output).

Now the below code ins included in our PDF, but I have opted not to have to see any of the warning messages that come with viewing it.

```
sex_codebook <-
  tibble(SEX = c(1,2),
         sex_clean = c("Male","Female"))

educ_codebook <-
  tibble(EDUC = c(0:22, 97:99),
         educ_clean = c(rep("No Degree", 14),
                        rep("HS Diploma",2),
                        rep("Some College", 3),
                        rep("College Degree", 4),
                        rep(NA, 3)
                        )
         )

slim_df <-
  df %>%
  select(AGE, SEX, EDUC, HEALTH, HEIGHT, WEIGHT) %>%
  sample_frac(.1, replace = F) %>%
  mutate(BMI = WEIGHT/HEIGHT) %>%
  left_join(sex_codebook, by = "SEX") %>%
  left_join(educ_codebook, by = "EDUC")
```

## Including tables

Let's say I now want to tell the story of how men and women get less healthy as they get older. The first thing I might want to show would be a table with the average health ratings of men and women compared with how old they are. I use the kable command to include a nice looking table in the markdown.

```
age_health_gender <-
  slim_df %>%
  group_by(AGE, sex_clean) %>%
  summarise(avg = mean(HEALTH)) %>%
  spread(key = sex_clean, value = avg)

kable(age_health_gender, caption = "Average health as respondents age by gender")
```

Table 1: Average health as respondents age by gender

| AGE | Female | Male |
|---|---|---|
| 0 | 1.552481 | 1.559778 |
| 1 | 1.585887 | 1.596570 |
| 2 | 1.658579 | 1.651990 |
| 3 | 1.609626 | 1.656327 |
| 4 | 1.631980 | 1.714523 |
| 5 | 1.678363 | 1.676589 |
| 6 | 1.660915 | 1.702352 |
| 7 | 1.699653 | 1.728856 |
| 8 | 1.665848 | 1.716654 |
| 9 | 1.673029 | 1.696302 |
| 10 | 1.722597 | 1.688872 |
| 11 | 1.689201 | 1.761062 |
| 12 | 1.653780 | 1.749390 |
| 13 | 1.692308 | 1.738115 |
| 14 | 1.732284 | 1.741087 |
| 15 | 1.772768 | 1.743232 |
| 16 | 1.752278 | 1.754596 |
| 17 | 1.797273 | 1.744367 |
| 18 | 1.864813 | 1.745045 |
| 19 | 1.909901 | 1.832139 |
| 20 | 1.936355 | 1.799819 |
| 21 | 1.939096 | 1.851378 |
| 22 | 1.946257 | 1.871694 |
| 23 | 1.912429 | 1.829848 |
| 24 | 1.942584 | 1.850049 |
| 25 | 1.967480 | 1.904939 |
| 26 | 1.933148 | 1.924138 |
| 27 | 1.957130 | 1.907772 |
| 28 | 2.009183 | 1.937198 |
| 29 | 1.909747 | 1.853586 |
| 30 | 2.080192 | 1.984375 |
| 31 | 2.033606 | 1.960915 |
| 32 | 1.989848 | 1.973684 |
| 33 | 2.086268 | 2.056550 |
| 34 | 2.087573 | 1.981273 |
| 35 | 2.102138 | 2.037866 |
| 36 | 2.099575 | 2.016870 |

| AGE | Female | Male |
|---|---|---|
| 37 | 2.128070 | 2.043311 |
| 38 | 2.154346 | 2.075343 |
| 39 | 2.122327 | 2.018617 |
| 40 | 2.188915 | 2.139283 |
| 41 | 2.172727 | 2.098729 |
| 42 | 2.247706 | 2.131695 |
| 43 | 2.236010 | 2.176311 |
| 44 | 2.245252 | 2.232108 |
| 45 | 2.296414 | 2.288642 |
| 46 | 2.325459 | 2.213894 |
| 47 | 2.332212 | 2.259908 |
| 48 | 2.334825 | 2.227586 |
| 49 | 2.392256 | 2.338332 |
| 50 | 2.402200 | 2.399667 |
| 51 | 2.399813 | 2.421158 |
| 52 | 2.512590 | 2.393004 |
| 53 | 2.530207 | 2.431316 |
| 54 | 2.427073 | 2.404914 |
| 55 | 2.523466 | 2.467807 |
| 56 | 2.486220 | 2.473738 |
| 57 | 2.543569 | 2.492257 |
| 58 | 2.573348 | 2.607399 |
| 59 | 2.534342 | 2.460208 |
| 60 | 2.635913 | 2.601198 |
| 61 | 2.626956 | 2.632411 |
| 62 | 2.663082 | 2.542933 |
| 63 | 2.656442 | 2.676796 |
| 64 | 2.639581 | 2.655930 |
| 65 | 2.603234 | 2.680384 |
| 66 | 2.630098 | 2.588907 |
| 67 | 2.716667 | 2.663158 |
| 68 | 2.688253 | 2.652651 |
| 69 | 2.717996 | 2.603516 |
| 70 | 2.702161 | 2.714286 |
| 71 | 2.715976 | 2.656751 |
| 72 | 2.780847 | 2.763797 |
| 73 | 2.851293 | 2.774193 |
| 74 | 2.882227 | 2.796748 |
| 75 | 2.848614 | 2.852547 |
| 76 | 2.850575 | 2.794030 |
| 77 | 2.935000 | 2.834835 |
| 78 | 2.933333 | 2.954693 |
| 79 | 2.982353 | 3.052083 |
| 80 | 2.964557 | 2.921933 |
| 81 | 2.893082 | 2.955357 |
| 82 | 2.872576 | 3.027523 |
| 83 | 3.129450 | 3.069519 |
| 84 | 2.966942 | 2.898477 |
| 85 | 3.105485 | 3.059603 |

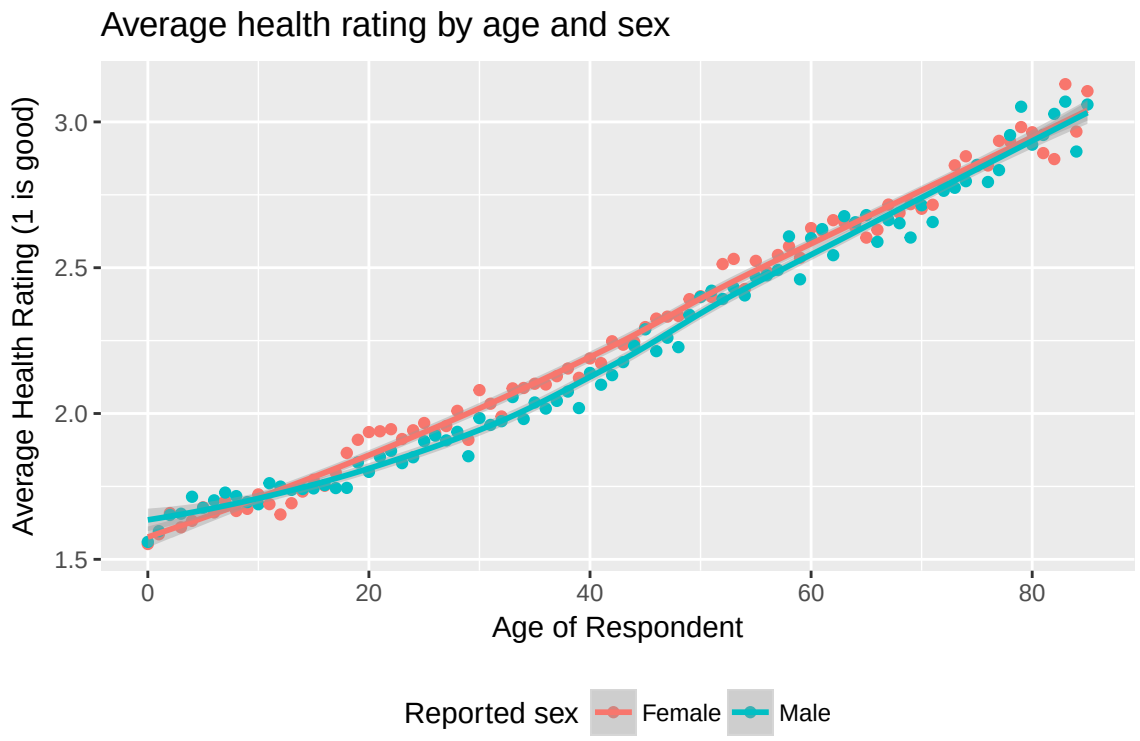
This table is so long it is useless, so let's represent this information differently.

## Including ggpplots

Now I want to further convince my readers by showing the difference in the distributions of our data for health rating by age and gender. To do this I will create a categorical age variable.

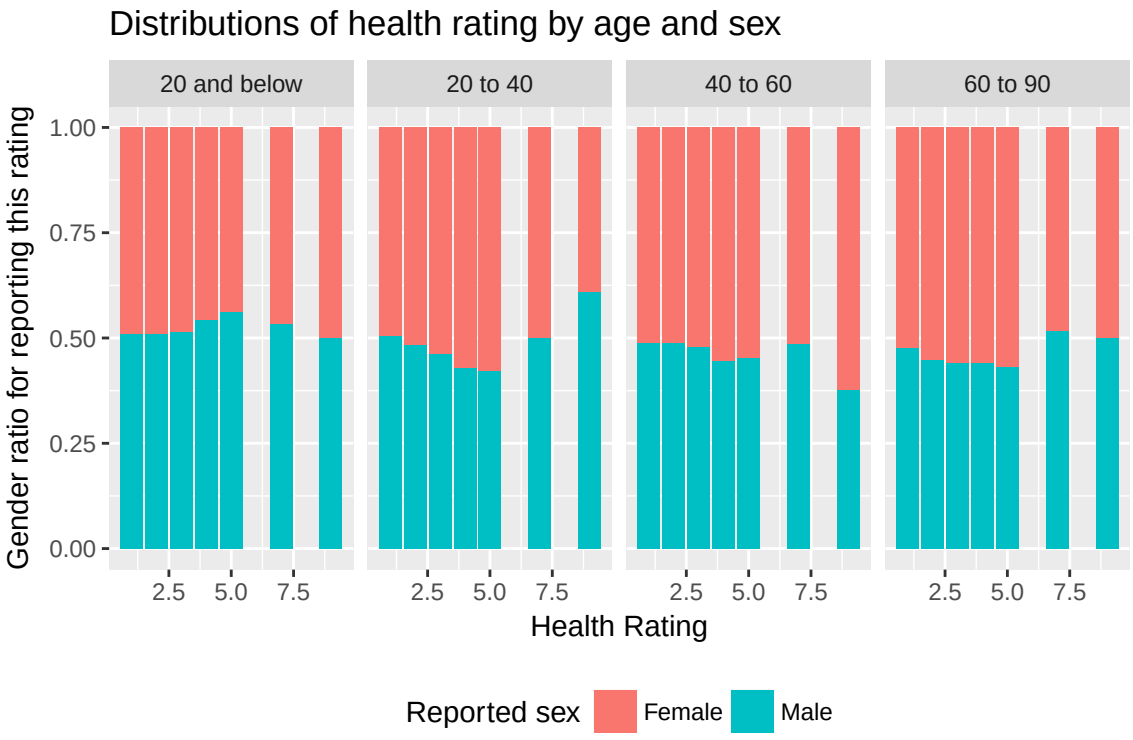
```
slim_df %>%
  group_by(AGE, sex_clean) %>%
  summarise(avg = mean(HEALTH)) %>%
  ggplot(aes(x = AGE, y = avg, color = sex_clean)) +
  geom_point() +
  geom_smooth() +
  theme(legend.position = 'bottom') +
  labs(title = "Average health rating by age and sex",
       x = "Age of Respondent",
       color = "Reported sex",
       y = "Average Health Rating (1 is good)")
```

```
## `geom_smooth()` using method = 'loess'
```




```
slim_df %>%
  mutate(age = case_when(AGE < 20 ~ "20 and below",
                         AGE < 40 ~ "20 to 40",
                         AGE < 60 ~ "40 to 60",
                         AGE < 90 ~ "60 to 90")) %>%
  ggplot(aes(x = HEALTH, fill = sex_clean)) +
  geom_bar(position = 'fill')+
  facet_grid(. ~ age) +
  theme(legend.position = 'bottom') +
  labs(title = "Distributions of health rating by age and sex",
       x = "Health Rating",
```

```
       fill = "Reported sex",
       y = "Gender ratio for reporting this rating")
```




## Including regression output

Now the below code chunk has a few options. The `results='asis'` option is there to make the output of stargazer render nicely. The `echo = F` prevents the source code form being inlcuded, and the `warning = F` prevents any warnings form being included in the PDF.

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sun, Feb 25, 2018 - 1:05:11 PM

Table 2:

| | *Dependent variable:* | |
|---|---|---|
| | HEALTH | |
| | (1) | (2) |
| sex_cleanMale | $-0.035^{***}$ | $-0.006$ |
| | (0.005) | (0.010) |
| AGE | $0.017^{***}$ | $0.018^{***}$ |
| | (0.0001) | (0.0002) |
| sex_cleanMale:AGE | | $-0.001^{***}$ |
| | | (0.0002) |
| Constant | $1.533^{***}$ | $1.519^{***}$ |
| | (0.005) | (0.007) |
| Observations | 161,170 | 161,170 |
| $R^2$ | 0.127 | 0.127 |
| Adjusted $R^2$ | 0.127 | 0.127 |
| Residual Std. Error | 1.013 (df = 161167) | 1.013 (df = 161166) |
| F Statistic | $11{,}735.190^{***}$ (df = 2; 161167) | $7{,}828.315^{***}$ (df = 3; 161166) |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |