Setting Up

We'll use the `midd_survey` data for this example. Load it and the usual packages.

Formatting Regression Tables With Stargazer

Summaries of linear regression models have a lot of information and can be messy in knitted files. One option to clean them up is to use the **stargazer** package. The package is highly customizable; we'll go over just a few options today. For more details about the package, click here.

To see how stargazer compares to the output we are used to, use the `midd_survey` data to regress `gpa` on `gender`, save the model as `model1`, and look at the summary (remember `echo = FALSE` suppresses the code in the knitted file):

```
##
## Call:
## lm(formula = gpa ~ siblings, data = midd_survey)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39967 -0.15695  0.04305  0.20487  0.60033
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.558767   0.017163 207.348  < 2e-16 ***
## siblings    -0.031819   0.009198  -3.459 0.000564 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3003 on 983 degrees of freedom
## Multiple R-squared:  0.01203,    Adjusted R-squared:  0.01102
## F-statistic: 11.97 on 1 and 983 DF,  p-value: 0.0005644
```

Now that you are comfortable interpreting summary outputs, you should continue using the traditional summary outputs to get a sense of your models. But replace `summary()` with `stargazer()` when you want to knit your file.

```
# INSTALL STARGAZER THE FIRST TIME FROM THE BOTTOM RIGHT PANE
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(model1)
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Mon, Dec 02, 2019 - 13:30:43
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
```

```
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & gpa \\
## \hline \\[-1.8ex]
##  siblings & $-$0.032$^{***}$ \\
##    & (0.009) \\
##    & \\
##  Constant & 3.559$^{***}$ \\
##    & (0.017) \\
##    & \\
## \hline \\[-1.8ex]
## Observations & 985 \\
## R$^{2}$ & 0.012 \\
## Adjusted R$^{2}$ & 0.011 \\
## Residual Std. Error & 0.300 (df = 983) \\
## F Statistic & 11.968$^{***}$ (df = 1; 983) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

There are a few other things to add to make sure the file knits properly:

- `results = 'asis'` is the most important thing. It leaves the output in the format that knitr needs to render it. (Add this option to the open fence line; add everything else to the stargazer function line.)
- `header = FALSE` deletes the contact information for the package author
- `type = 'latex'` converts the output into the Latex language that pandoc needs to knit to a PDF. Latex is the default so you can leave it off if you want. To change to web output you would want to change `type = 'latex'` to `type = 'html'`.

Table 1:

	Dependent variable:
	gpa
siblings	−0.032***
	(0.009)
Constant	3.559***
	(0.017)
Observations	985
R^2	0.012
Adjusted R^2	0.011
Residual Std. Error	0.300 (df = 983)
F Statistic	11.968*** (df = 1; 983)
Note:	*p<0.1; **p<0.05; ***p<0.01

Knit the file to see the formatted regression table.

This is better than dumping raw R output into a report, but we can do even better.

Here are some more changes for the `stargazer()` function:

- `style = 'ajs'` formats the table following publication guidelines for the *American Journal of Sociology*. Replace 'ajs' with 'asr' to use the formatting style for the *American Sociological Review*. These are the two most common formatting styles you will see for sociology papers. I prefer ajs style for written reports (because it includes standard errors), and asr style for presentation slides (because it is more condensed).
- `title` creates a title for the table which will be numbered automatically
- `covariate.labels = c()` is a list of names for the independent and control variables in your model. You need to label every variable or this won't work.
- `dep.var.labels =` creates a title above the dependent variable model numbers
- `keep.stat = c("n", "rsq")` reports the number of observations and the r-squared value but excludes all the other model information

We can include more than one model in a table too. Let's save a model with a control variable and a model with an interaction.

And then include our three models in a stargazer function that also includes several of the changes described above.

```
stargazer(model1, model2, model3,
          header=FALSE, type = 'latex', style = 'asr',
  title = "A Better Title",
  covariate.labels = c("Number of Siblings", "Gender = Other", "Gender = Woman",
                       "Siblings X Gender = Other", "Siblings X Gender = Woman"),
  dep.var.labels = "GPA",
  keep.stat = c("n", "rsq"))
```

Table 2: A Better Title

	GPA		
	(1)	(2)	(3)
Number of Siblings	-0.032^{***}	-0.031^{***}	-0.023
Gender = Other		-0.058	0.296^{*}
Gender = Woman		0.046^{*}	0.056
Siblings X Gender = Other			-0.269^{***}
Siblings X Gender = Woman			-0.006
Constant	3.559^{***}	3.532^{***}	3.519^{***}
N	985	985	985
R^2	0.012	0.018	0.033

*p $< .05$; **p $< .01$; ***p $< .001$