# Notes on the Data Set

The data set used in these exercises is a subset of the Pew Religious Landscape Survey conducted in 2014 by the Pew Research Center. The survey was a very large telephone survey of about 35,000 adults living in all 50 states in the U.S. About 60% of all interviews were on cell phones and 40% on landlines. At least 300 interviews were conducted in each state including the District of Columbia. More information about the survey can be found on the [Pew website](http://www.pewforum.org/religious-landscape-study/). You will find their website extremely useful. A similar survey of about 35,000 adults was conducted by the Pew Center in 2007.

## Selecting the Variables

A subset of variables was selected from the full Pew survey. Some of these variables were created by combining other variables[[1]](#footnote-1) while other variables were recodes of existing variables. Variables were selected that would be most useful for instructional purposes. The full data set can be downloaded from the [Pew Center's website](http://www.pewforum.org/datasets/). Be sure to read the Dataset Use Agreement before downloading the data set.

Weighting

There is a weight variable that should be used whenever analyzing the data. Weighting adjusts the data so it better represents the population from which the sample was selected.

The survey also includes another set of weights that should be used if you want to focus on specific metropolitan areas. These weights are **not** included in this subset. If you want to look at certain metropolitan areas you will need to download the full data set from the Pew website.

## Renaming Variables

Variables were renamed to make them easier to use. Names consist of a short description of the type of variable and a unique numbering of the variables. Variable names are kept short for convenience sake. Here is a listing of the different types of variables in the dataset.

* AB1 – abortion
* C1 to C3 – various types of changes in the U.S. and whether this was for the better or worse
* D1 to D14 – demographic variables including marital status, race, ethnicity, children, age, cohort, education, citizen, family income, and sex
* ENV1 – environmental laws and regulation
* EVO1 – evolution
* G1 to G2 – geographic region including state and census region
* GA1 – government aid to the poor
* GS1 – size of government
* H1 to H2 – homosexuality and whether they know anyone who is gay or lesbian
* HS1 to HS3 – satisfaction with different areas of life
* ID – unique identification number
* P1 to P5 – political party identification and ideology
* R1 to R7 – religious identification
	+ present religion
	+ born-again or evangelical Christian
	+ religious family including evangelical Protestant tradition, mainline Protestant tradition, historically black Protestant tradition, Catholic, Jewish, Muslim, Buddhist, Hindu, other, atheist, agnostic, nothing in particular; the various Protestant traditions are broken down into their respective denominations (e.g., Lutheran, Methodist)
	+ same as above but Protestant traditions are not broken down into their respective denominations
	+ Protestant denominations with non-Protestants coded as not Protestant
* REL1 to REL3 – religiosity including attendance at worship, importance of religion, and frequency of prayer
* RR1 to RR4 – religion in which raised
* RS1 to RS4 – religion of spouse or partner
* RBH1 to RBH11 – religious behavior including such things as participation in small religious groups and reading of scripture
* RBL1 to RBL21 – religious beliefs including such things as belief in God, view of God, and existence of heaven and hell
* RW1 – what respondents look to for guidance on right and wrong
* STAN1 – absolute standards of right and wrong
* SS1 – same-sex marriage
* WEIGHT variable

## Variable Labels

Variable labels always start with the variable name. The rest of the label is the label that Pew used in the full data set. This identifies the question number in the survey and the wording of each question. This is very useful since you can see exactly how the question was asked in the survey. I did some minor editing of the Pew variable labels to make them easier to read.

## Composite Variables

Some variables were created by Pew and some I created. These are identified in the variable label as a "composite or a combined variable." Pew created the following variables: R4, R5, R6, RR2, RR3, RR4, RS2, RS3, and RS4. I created the following variables: D10, EVO1, RBH7, P4, and RBL21.

The first three of the variables I created were straight forward combinations of two other variables.

* D10 combines two family income questions into one variable.
* EVO1 combines two questions about evolution into one variable.
* RBH7 combines two questions about volunteering into one variable.

The last two variables I created (P4 and RBL21) requires a little more explanation.

* P2 asked respondents whether they considered themselves to be Republican, Democrat, or Independent. Some respondents added that they had no preference or belonged to another party.
* P3 asked those that responded independent, no preference, belonged to another party, and those that said they didn't know or refused to answer another question that asked whether they leaned more to the Republican or Democratic Party.
* I combined these two variables into another variable (P4). This variable classifies respondents as Republican, lean Republican, Independent, lean Democrat, and Democrat. Independents consist of those who said they were Independent in P2 and didn't lean either way in P3.
* Some of the exercises look at two defining beliefs of Evangelical Christians – the belief that one must have had a born-again experience (R3) and the belief that the Bible is the literal word of God (RBL7R2). RBL21 combines these two beliefs into a typology of Christian beliefs. Your analysis will automatically be limited to Christians since non-Christians are defined as missing data.

## Recoded Variables

Several variables were recoded to make them easier to use. These include RBL7R1, RBL7R2, D6R1, D7R2, D8R1, and D10R1. The R toward the end of the variable name indicates that it is a recoded variable and the number at the very end of the name distinguishes between the first and second recodes.

## Missing Values

I created missing values for all variables that had any missing information. I used the following rules to create the missing values.

* Don't know and refused were always defined as missing values.
* In variables C1 to C3 respondents were asked if various changes were for the better or for the worse. Some said that the changes were mixed and I assigned them a missing value.
* Several variables were forced choice questions where the respondents were asked to choose the statement that was closest to their view. Some said both or neither and I assigned them a missing value.
* Some responses were uncodeable and I assigned them a missing value.

## Levels of Measurement

Variables are often labelled as nominal, ordinal, interval, and ratio level based on a classification scheme developed by S. S. Stevens. The variables in this subset are all nominal or ordinal level measurement. There are no interval or ratio level measures. This, of course, limits the type of statistical analysis you can do with these variables. If you want a series of exercises that include interval and ratio level variables, I created three sets of exercises using the General Social Survey. These use SPSS, PSPP, and SDA (Survey Documentation and Analysis) as the statistical package for analysis. You can find these exercises on the [Social Science Research and Instructional Council's website](http://ssric.org/trd/exercises).

## Statistical Package Used

I use SPSS as the statistical package for these exercises. There are other statistical packages that you might use including SAS, Stata, R, and others. You can convert the data set to other formats as you wish.

1. Some of these variables were created by Pew and some I created. [↑](#footnote-ref-1)