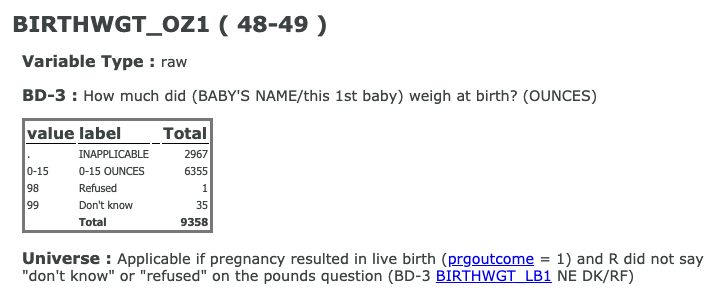
**Read the codebook**

When you work with datasets like the NSFG, it is important to read the documentation carefully. If you interpret a variable incorrectly, you can generate nonsense results and never realize it. So, before we start coding, I want to make sure you are familiar with the NSFG codebook, which describes every variable.

* Follow [**this link**](https://www.icpsr.umich.edu/icpsradmin/nsfg/index?studyNumber=9999) to get to the interactive codebook.
* Type "birthweight" in the search field, UNSELECT the checkbox that says "Search variable name only", and press "Search". You should see a list of variables related to birthweight.
* Click on "BIRTHWGT\_OZ1" and read the documentation of this variable. For your convenience, it is also displayed here:



How many respondents refused to answer this question?

**Answer the question**

**50 XP**

**Possible Answers**

1

35

48-49

2967

**Exploring the NSFG data**

To get the number of rows and columns in a DataFrame, you can read its shape attribute.

To get the column names, you can read the columns attribute. The result is an Index, which is a Pandas data structure that is similar to a list. Let's begin exploring the NSFG data! It has been pre-loaded for you into a DataFrame called nsfg.

**Instructions 1/4**

**25 XP**

* Calculate the number of rows and columns in the DataFrame nsfg.

Display the names of the columns in nsfg.

Select the column 'birthwgt\_oz1' and assign it to a new variable called ounces.

Display the first 5 elements of ounces.

In [1]: # Display the number of rows and columns

nsfg.shape

nsfg.columns

Out[1]: Index(['caseid', 'outcome', 'birthwgt\_lb1', 'birthwgt\_oz1', 'prglngth', 'nbrnaliv', 'agecon', 'agepreg', 'hpagelb', 'wgt2013\_2015'], dtype='object')

In [2]: # Display the number of rows and columns

nsfg.shape, nsfg.columns

Out[2]:

((9358, 10),

Index(['caseid', 'outcome', 'birthwgt\_lb1', 'birthwgt\_oz1', 'prglngth', 'nbrnaliv', 'agecon', 'agepreg', 'hpagelb', 'wgt2013\_2015'], dtype='object'))

<script.py> output:

0 4.0

1 12.0

2 4.0

3 NaN

4 13.0

Name: birthwgt\_oz1, dtype: float64

In [3]: