Overview:

The SWAN data files can be downloaded in multiple formats. The lowest incidence of error was seen in R Studio, so RDA files were used to combine data, then later saved into a .csv filetype so that R and python can both be leveraged for this project.

Files:

Table 1 : raw SWAN data files. (filepath: SWAN_capstone_delivery --> SWAN_Raw_Data)

Filename	SWAN visit number	Year(s) of visit
28762-0001-Data.rda	Baseline	1997
29221-0001-Data.rda	Visit 1	1997 -1999
29401-0001-Data.rda	Visit 2	1998 - 2000
29701-0001-Data.rda	Visit 3	1999 - 2001
30142-0001-Data.rda	Visit 4	2000 - 2002
30501-0001-Data.rda	Visit 5	2001 - 2003
31181-0001-Data.rda	Visit 6	2002 - 2004
31901-0001-Data.rda	Visit 7	2003 - 2005
32122-0001-Data.rda	Visit 8	2004-2006
32721-0001-Data.rda	Visit 9	2005 -2007
32961-0001-Data.rda	Visit 10	2006 - 2008

Table 2: Data Files created for this study: (filepath: SWAN_capstone_delivery --> Clean)

Filename	Filetype	Description	Dependencies
combining_data.R	R coding script	Code used to combine and clean SWAN .rda data	All files in table 1
full_data_long.rda	R data file	Cleaned and joined SWAN data, saved in R data formatting	combining_data.R
full_data_long.csv	CSV data file	Cleaned and joined SWAN data, saved in CSV data formatting	combining_data.R
cleanData.csv	CSV data file	The resulting file after the main portion of modeling was complete.	Capstone_modeling.ipynb

Table 3: SWAN Analyses: (filepath: SWAN_capstone_delivery --> SWAN_Analyses)

Filename	Filetype	Description	Dependencies
Capstone_modeling.ipynb	Jupityr Notebook	Primary modeling work - sleep index created	full_data_long.csv
Mixed_Effects_Model.R	R coding script	Modeling nutritional and supplement data against sleep index.	full_data_long.csv
motherhood_model.ipynb	Jupityr Notebook	Modeling motherhood indicators against sleep index	full_data_long.csv
Supps_Labs_EDA.ipynb	Jupityr Notebook	Exploration of supplements and nutritional markers.	full_data_long.csv
exercise.ipynb	Jupityr Notebook	Modeling for exercise's influence on sleep quality	30142-0001-Data.tsv (this is the visit 4 data in a tab separated value file, download at SWAN). This file is used because the 4th visit is the only one where all three exercise types exist.