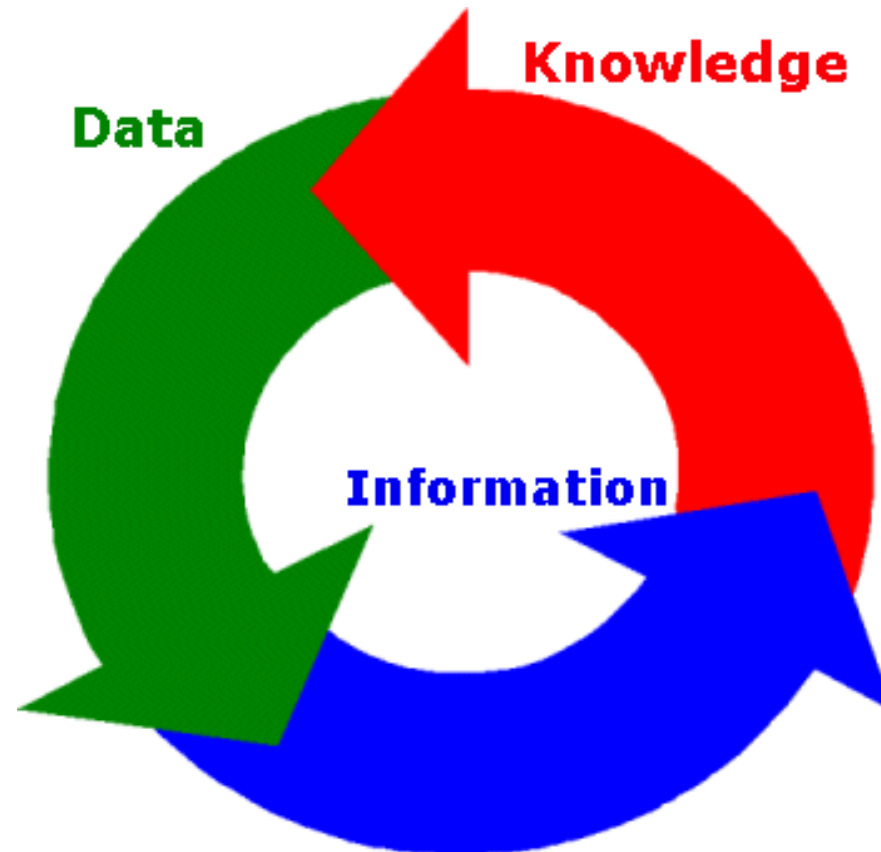
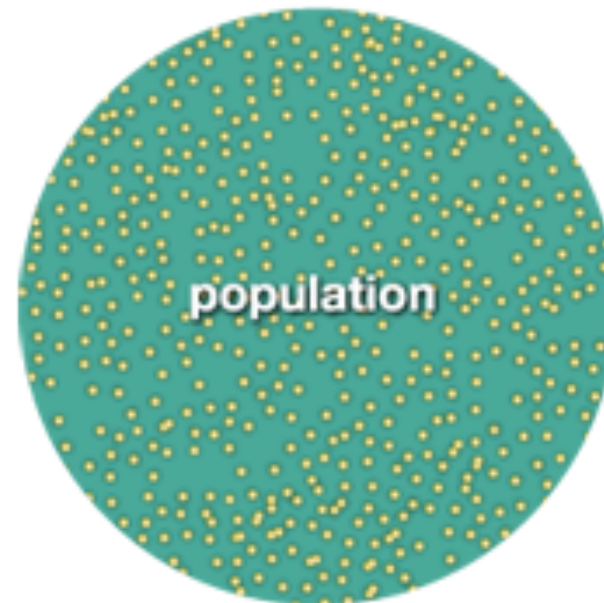


What is statistics?









Height
Weight
Gender
Race



Color
Mileage
Model
of seats



Population size
Tax revenue
Energy consumption
of hospitals



Variables

Individuals		Gender (M/F)	Age	Weight (lbs.)	Height (in.)	Smoking (0=No, 1=Yes)	Race
	Patient #1	M	59	175	69	0	White
	Patient #2	F	67	140	62	1	Black
	Patient #3	F	73	155	59	0	Asian

	Patient #75	M	48	190	72	0	White

Quantitative=Q

Categorical=C



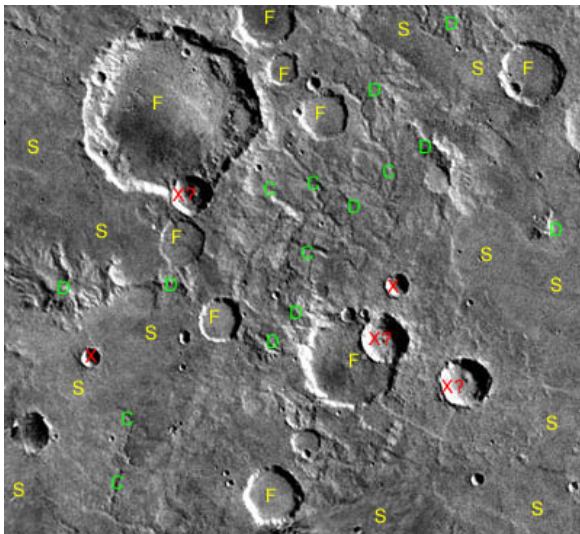
Open Learning Initiative

Transforming higher education through the science of learning.

Available Data Sets



The National Longitudinal Study
of Adolescent Health
(AddHealth)



Mars Craters Study
<http://craters.sjrdesign.net>

Available Data Sets



The National Epidemiologic Study of Alcohol and Related Conditions (NESARC)



Gap Minder Study

<http://www.gapminder.org>

Example from AddHealth Code Book

*In Home Questionnaire Code Book, S.1
Public Use Sample*

Frequency	Code	Response	Variable Name	Type/ Length
IMPORTANT: See NOTE at beginning of codebook.				
20.		What grade {ARE/WERE} you in? <i>[If school doesn't have grade levels of this kind, enter "99."]</i>	H1GI20	num 2
979	7	seventh grade		
992	8	eighth grade		
1107	9	ninth grade		
1144	10	tenth grade		
1122	11	eleventh grade		
993	12	twelfth grade		
1	96	refused		
128	97	legitimate skip <i>[not in school]</i>		
3	98	don't know		
35	99	school doesn't have grade levels of this kind or not applicable		

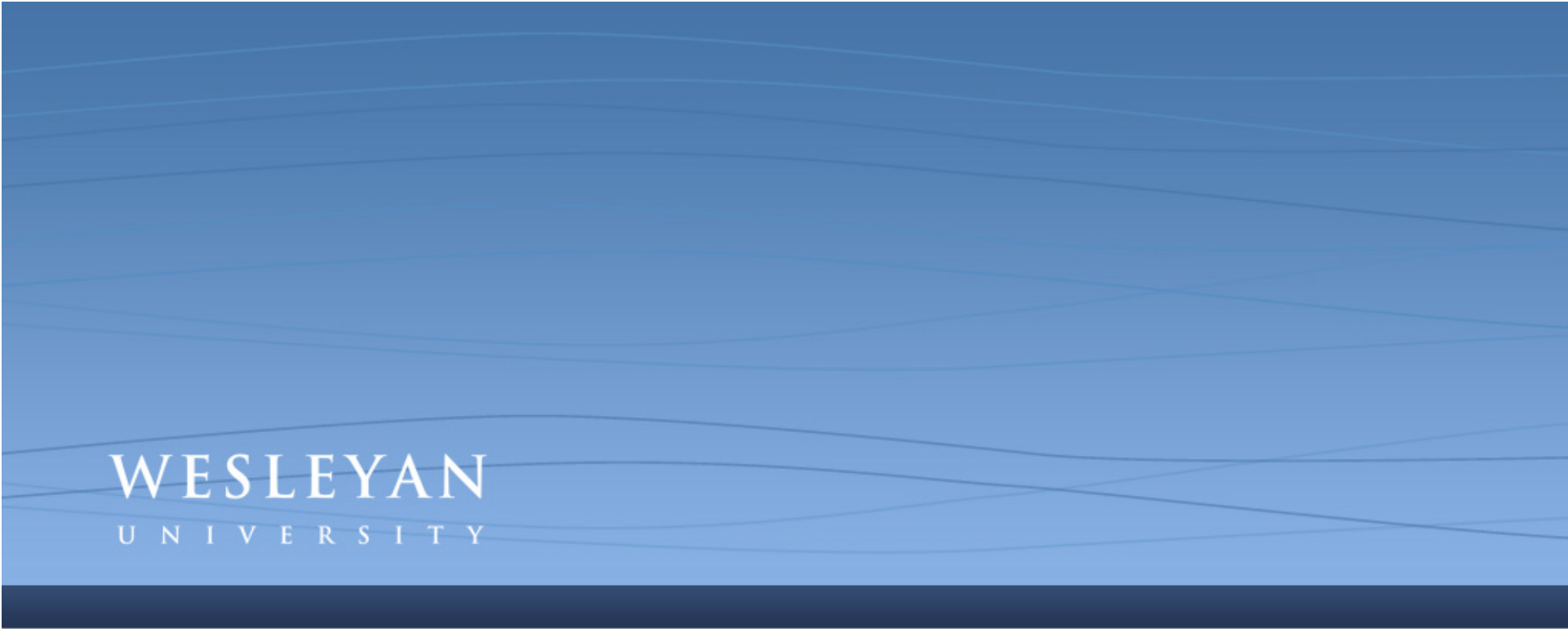
Example from Mars Craters Code Book

Variable Names with Descriptions

- CRATER_NAME – crater unique identifier, based upon the region of the planet (1/16ths), the “pass” under which the crate was identified, and the order in which it was identified
- LATITUDE_CIRCLE_IMAGE – latitude from the derived center of a non-linear least-squares circle fit to the vertices selected to manually identify the crater rim (units are decimal degrees North)
- LONGITUDE_CIRCLE_IMAGE – longitude from the derived center of a non-linear least-squares circle fit to the vertices selected to manually identify the crater rim (units are decimal degrees East)
- DIAM_CIRCLE_IMAGE – diameter from a non-linear least squares circle fit to the vertices selected to manually identify the crater rim (units are km)
- DEPTH_RIMFLOOR_TOPOG – average elevation of each of the manually determined N points along (or inside) the crater rim (units are km)
 - Depth Rim - Points are selected as relative topographic highs under the assumption they are the least eroded so most original points along the rim
 - Depth Floor – Points were chosen as the lowest elevation that did not include visible embedded craters

Example from GapMinder Code Book

Variable Name	Description of Indicator	Main Source
incomeperperson	2010 Gross Domestic Product per capita in constant 2000 US\$. The inflation but not the differences in the cost of living between countries has been taken into account.	World Bank Work Development Indicators
alcoholconsumption	2008 alcohol consumption per adult (age 15+), litres Recorded and estimated average alcohol consumption, adult (15+) per capita consumption in litres pure alcohol	WHO
armedforcesrate	Armed forces personnel (% of total labor force)	Work Development Indicators
breastcancerper100 TH	2002 breast cancer new cases per 100,000 female Number of new cases of breast cancer in 100,000 female residents during the certain year.	ARC (International Agency for Research on Cancer)

The image features a solid blue background with several thin, light blue wavy lines that create a sense of movement. In the lower-left corner, the Wesleyan University logo is displayed in white. The word "WESLEYAN" is in a large, serif font, and "UNIVERSITY" is in a smaller, sans-serif font directly below it.

WESLEYAN
UNIVERSITY

A word cloud where the words 'ASSIGNMENT' and 'ONE' are the primary focus. 'ASSIGNMENT' is written in a large, bold, blue font, appearing multiple times in different orientations. 'ONE' is written in a smaller, blue font, also appearing multiple times. The background is white, and the words are scattered across the frame, with some overlapping. The overall effect is a dense, repetitive visual of the two words.

Select a topic of interest

3650-3650	TAB12MDX	NICOTINE DEPENDENCE IN THE LAST 12 MONTHS
	38131	0. No nicotine dependence
	4962	1. Nicotine dependence
3652-3652	TABLIFEDX	NICOTINE DEPENDENCE - LIFETIME
	36156	0. No nicotine dependence
	6937	1. Nicotine dependence
3653-3653	STIM12ABDEP	AMPHETAMINE ABUSE/DEPENDENCE IN LAST 12 MONTHS
	43032	0. No amphetamine diagnosis
	34	1. Amphetamine abuse only
	3	2. Amphetamine dependence only
	24	3. Amphetamine abuse and dependence
3654-3654	STIMP12ABDEP	AMPHETAMINE ABUSE/DEPENDENCE PRIOR TO THE LAST 12 MONTHS
	42342	0. No amphetamine diagnosis
	530	1. Amphetamine abuse only
	25	2. Amphetamine dependence only
	196	3. Amphetamine abuse and dependence

Is medical treatment seeking associated with income?

Is humidity associated with caterpillar reproduction?

Is water fluorination associated with number of cavities during dentist visits?

Select a second topic

721-721	CHECK321	CIGARETTE SMOKING STATUS	

		9913	1. Smoked cigarettes in the past 12 months
		8078	2. Smoked cigarettes prior to the last 12 months
		22	9. Unknown
		25080	BL. NA, never or unknown if ever smoked 100+ cigarettes

722-722	S3AQ3B1	USUAL FREQUENCY WHEN SMOKED CIGARETTES	

		14836	1. Every day
		460	2. 5 to 6 Day(s) a week
		687	3. 3 to 4 Day(s) a week
		747	4. 1 to 2 Day(s) a week
		409	5. 2 to 3 Day(s) a month
		772	6. Once a month or less
		102	9. Unknown
		25080	BL. NA, never or unknown if ever smoked 100+ cigarettes

723-724	S3AQ3C1	USUAL QUANTITY WHEN SMOKED CIGARETTES	

		17751	1-98. Cigarette(s)
		262	99. Unknown
		25080	BL. NA, never or unknown if ever smoked 100+ cigarettes