Datasets
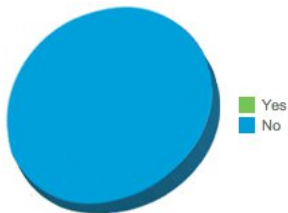
Variables

- Quantitative data is stored as variables of different types.
- A variable is something that varies over $k > 1$ states.
- Variables follow different scales and standards of measurement.
- Variable conform to a particular coding to allow for storage.

$k = 1$ (constant)

Values and labels

- **Values:** -1 0.7 9 (numeric), "UK" (string), . (missing)
- **Variable labels:** "Fertility rate"
- **Value labels:** 1 "White" 2 "Black" 3 "Hispanic" ...

Dimensions   Handbook

- **Continuous**: infinite/large number of possible values
- **Categorical**: finite/small number of possible values
- **Orders**: nominal (no possible order), ordinal, interval, ...

Basic principles

Mathematical ideals

- **Continuous random variables** can take any value on a given segment of the real number line $\mathbb{R}$.
- **Discrete random variables** take a finite number of possible values $X = x_1, x_2, ..., x_k$, with $\sum Pr(X) = 1$.

Social science perspective

- **Quantitative issue:** accuracy of the measurement
- **Qualitative issue**: meaningfulness of the unit

Examples: unit-less indexes, baseline growth rates, 'standardized' Likert scales, breaks in methodology and/or availability, . . .

Recoding

Rationale

- **Create groups**, e.g. age cohorts, social classes
- **Create dummies**, i.e. binary (0/1), true/false indicators
- **Change encoding** for missing values

Stata commands   Stata Guide, Sections 5–8

- `gen`, `tab`, `gen()`, `clonevar`
- `recode`, `irecode`, `replace`...
- `la var`, `la def`, `la val`...

Practice: WVS dataset

"[The government] should implement only the laws of the sharia."

- Measured on a 5-pt scale of agreement
- Measured in 9 African and Asian countries
- Measured between 1999 and 2004

Data:

- World Values Survey (WVS)
- Sample: resident populations aged 15+

Practice session

Class

```
* Get the demo code.
srqm fetch week3.do
* Open the do-file for this week.
doedit code/week3
```

Coursework

- Finish the do-file and read all comments at home.
- Follow instructions on top of the code.
- Confirm your group by next week, and choose a dataset.
- Select some continuous and categorical variables for analysis.

Exercises

Ex 3.1. European Social Survey 2008

1. Load the data.
2. Find the variable measuring left/right political positioning.
3. How continuous is the measurement?

Ex. 3.2. Quality of Government 2011

1. Load the data.
2. Find all variables on democracy.
3. How has democracy been measured?
   (Read the codebook too.)