Strategic Data Science (SDS)

# Data Methods

Karl Ho
School of Economic, Political and Policy Sciences
University of Texas at Dallas

# Day 2 Outline:

I. Data Collection Methods

1. Made data
2. Found data

II. Programming Workshop

1. R with Rstudio

# Data Methods

# Data Methods

1. Survey

# Data Methods

1. Survey
2. Experiments

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data
7. Complex Data
    1. Network Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data
7. Complex Data
    1. Network Data
    2. Multiple-source linked Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data
7. Complex Data
    1. Network Data
    2. Multiple-source linked Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data
7. Complex Data
    1. Network Data
    2. Multiple-source linked Data

Made Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data
7. Complex Data
    1. Network Data
    2. Multiple-source linked Data

Made Data

# Data Methods

1. Survey
2. Experiments
3. Qualitative Data
4. Text Data
5. Web Data
6. Machine Data
7. Complex Data
    1. Network Data
    2. Multiple-source linked Data

Made Data

Found Data

# What is survey?

American Statistical Society (ASA)

”Survey" is used most often to describe a method of gathering information from a sample of individuals. This "sample" is usually just a fraction of the population being studied.

# Illustration: HKES

YouGov has conducted two waves of election surveys in 2016 before and after the Legislative election. The company provided multiple weights created using rim weighting (also called Raking) using the following data:

1. Registered voter gender
2. Registered voter age
3. Registered voter district
4. Education based on Pre-election survey result
5. Income based on Pre-election survey result

# Illustration: HKES

# Illustration: HKES

The pre and post weights have maximum values to 18.

# Illustration: HKES

The pre and post weights have maximum values to 18.

The general weight value is under 5.

# Illustration: HKES

The pre and post weights have maximum values to 18.

The general weight value is under 5.

# Illustration: HKES

Possible reasons were:

# Illustration: HKES

Possible reasons were:

1. Weights were created using different populations

# Illustration: HKES

Possible reasons were:

1. Weights were created using different populations
2. Panelists were more representative of the younger population

# Illustration: HKES

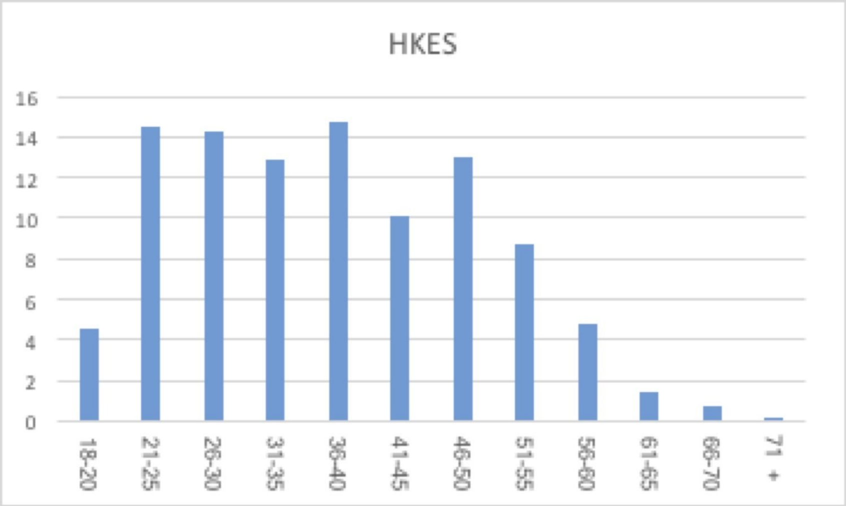
# Illustration: HKES

For Point 1:

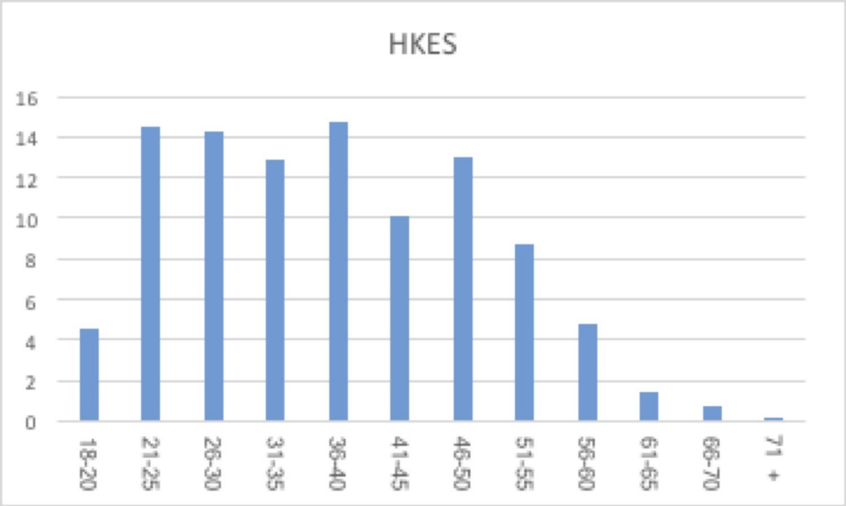
Hong Kong population has a male to female ratio of 47:53 according to the Census.  Registered voter population however has an even distribution of *49:51.*

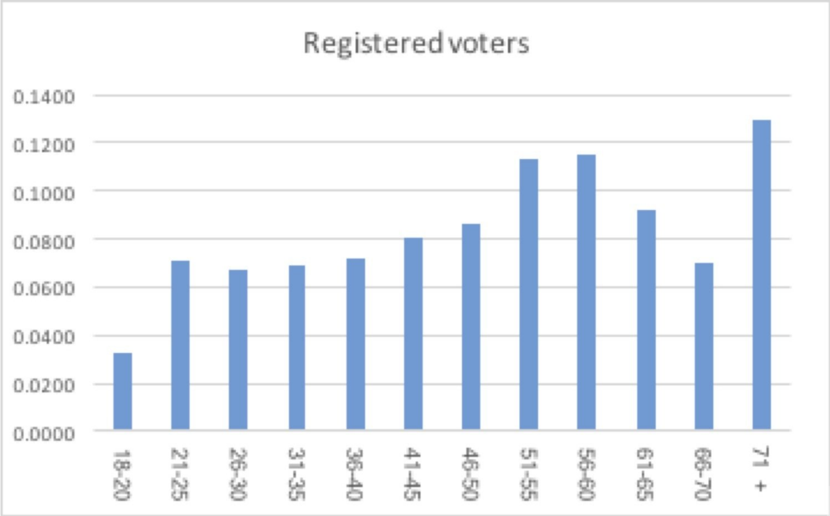
# Illustration: HKES

# Illustration: HKES



# Illustration: HKES



# Illustration: HKES

# Illustration: HKES

Previous figures illustrate the big difference between the HKES sample (which has more panelists from the younger group) and registered voter population. The latter indicates a large proportion in the elderly population. This can be attributed to some political parties' concerted efforts in mobilizing the elderly to register to vote.

# Illustration: HKES

Source: SCMP http://www.scmp.com/news/hong-kong/article/1855887/hong-kong-elderly-sign-droves-vote-district-council-elections

# Illustration: HKES

# Illustration: HKES

# Illustration: HKES

For point 2, YouGov acknowledges that the company has more access to the younger population via their recruitment channel.  It can be due to the highly savvy and active internet user population in the younger age groups.

# Illustration: HKES

For point 2, YouGov acknowledges that the company has more access to the younger population via their recruitment channel. It can be due to the highly savvy and active internet user population in the younger age groups.

Another reason that can be posting a problem is using two other demographic variables education and income from other population, that can be more representative of the population or the online population but not necessary the registered voter population.

# Illustration: HKES

# Illustration: HKES

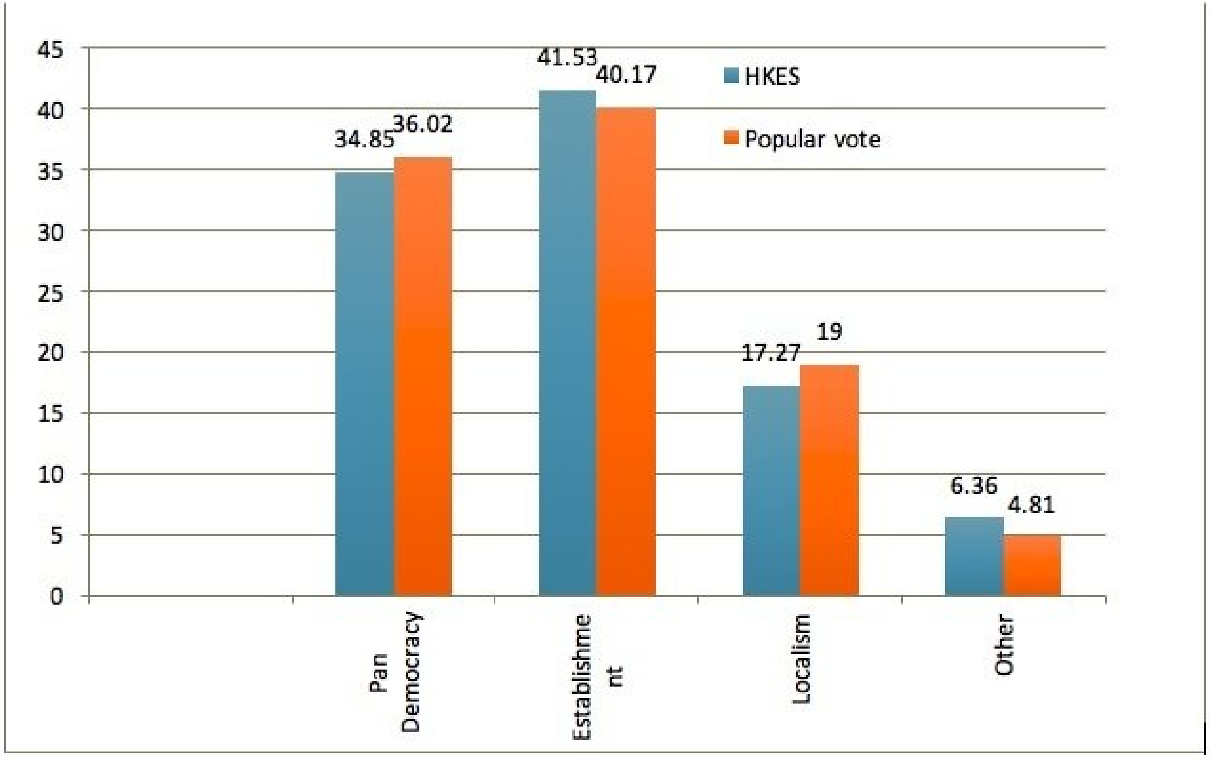
Raking is employed to generate a weight using age, gender and district only. The range of the weight for pre wave is from .269 to 8.939. They are slightly less varied that the original weights.

# Illustration: HKES

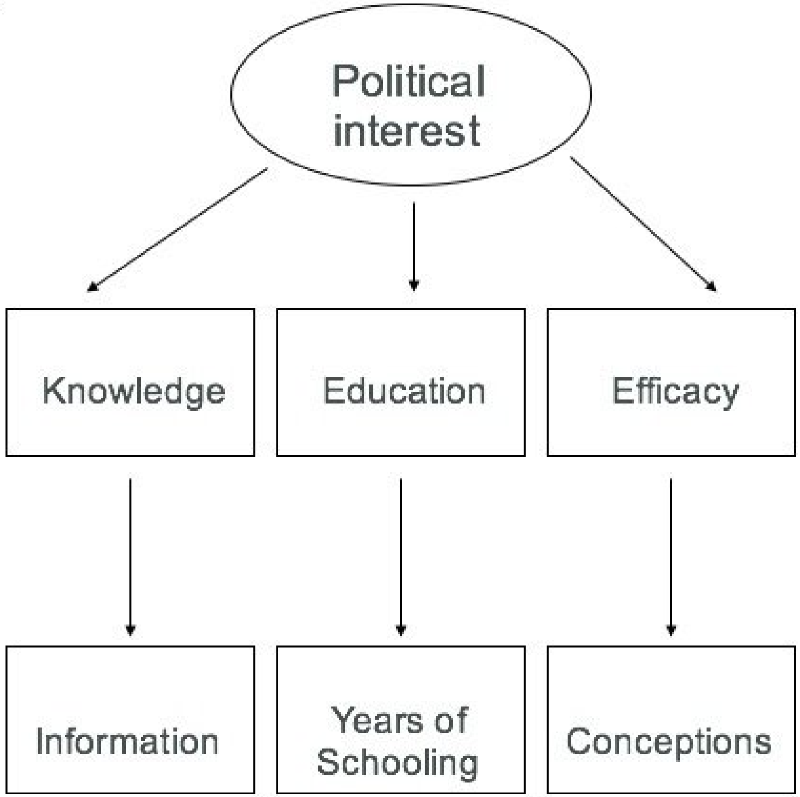
Raking is employed to generate a weight using age, gender and district only. The range of the weight for pre wave is from .269 to 8.939. They are slightly less varied that the original weights.

# Illustration: HKES

Figure 2. HKES vote camp support and actual popular votes



# Measurement




**Construct:**
abstract, theoretical, hypothetical
can't observe/measure

**Variable:** *reflects* construct, but is directly measurable and can differ from subject to subject (not a *constant*). Variables can be *Discrete or Continuous.*

**Operational Definition:**
concrete, measurable

*Defines variable by specific operations used to measure it*

# Data Methods: Experiments

Karl Ho
School of Economic, Political and Policy Sciences
University of Texas at Dallas

We are limited by the impossibility of experiment. Politics is an observational, not an experimental science . . . ”
(Lawrence Lowell, APSA President 1910)

# Why experimental design?

# Why experimental design?

Experiments facilitate causal inference through the transparency and content of experimental procedures, most notably the random assignment of observations to control and treatment groups.

# Why experimental design?

Experiments facilitate causal inference through the transparency and content of experimental procedures, most notably the random assignment of observations to control and treatment groups.

- Druckman et al. 2006

# What was the term "experiment" used in APSR?

# What was the term "experiment" used in APSR?

1. an institutional innovation such as a new constitution, electoral system, or policy process

# What was the term "experiment" used in APSR?

1. an institutional innovation such as a new constitution, electoral system, or policy process
2. a simulation or an empirical test that involves neither an institution nor randomized trials

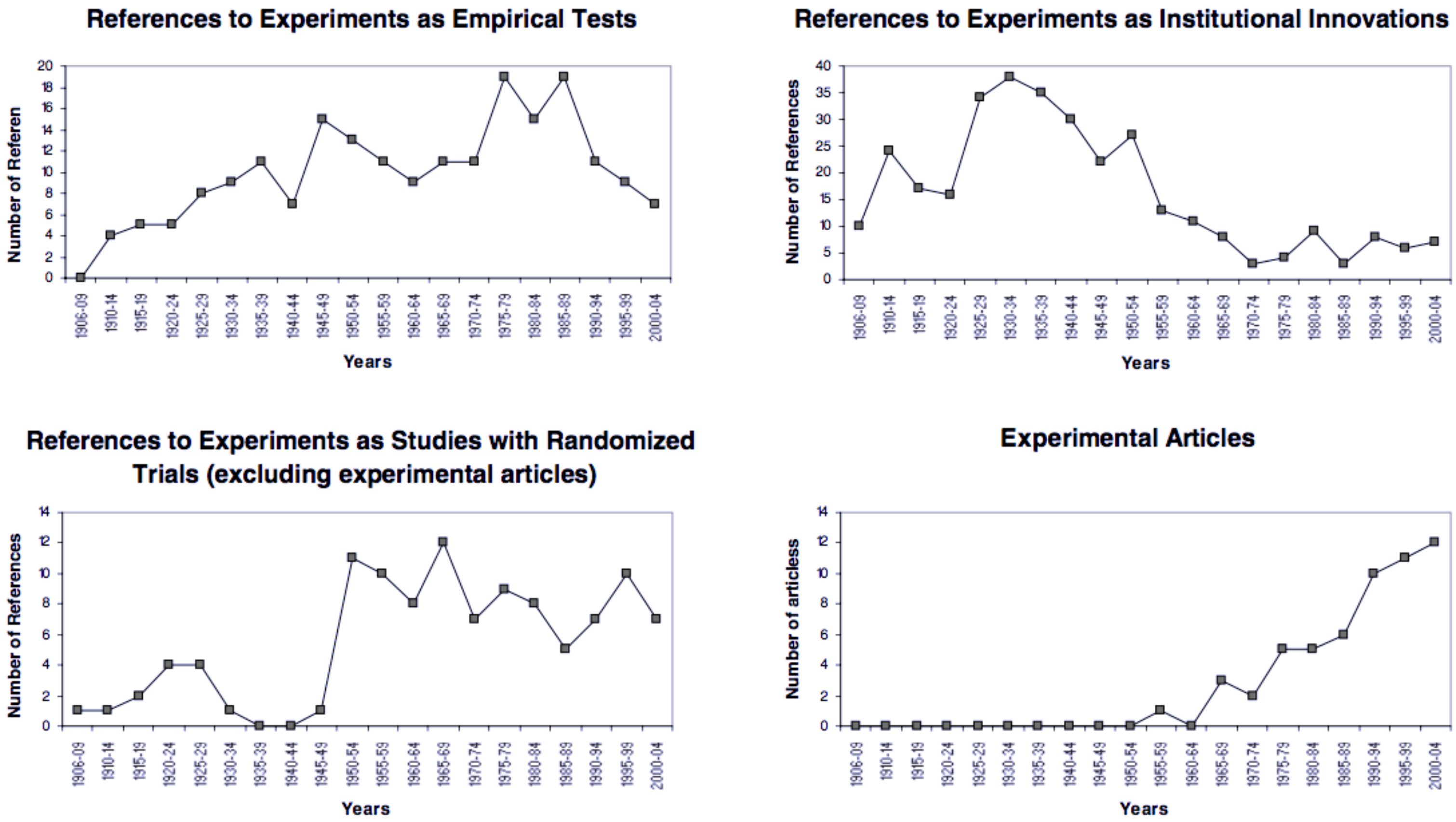
# What was the term "experiment" used in APSR?

1. an institutional innovation such as a new constitution, electoral system, or policy process
2. a simulation or an empirical test that involves neither an institution nor randomized trials
3. a randomized trial in which the researcher randomly assigns units of observation to control and treatment groups.

# Why experimental design?

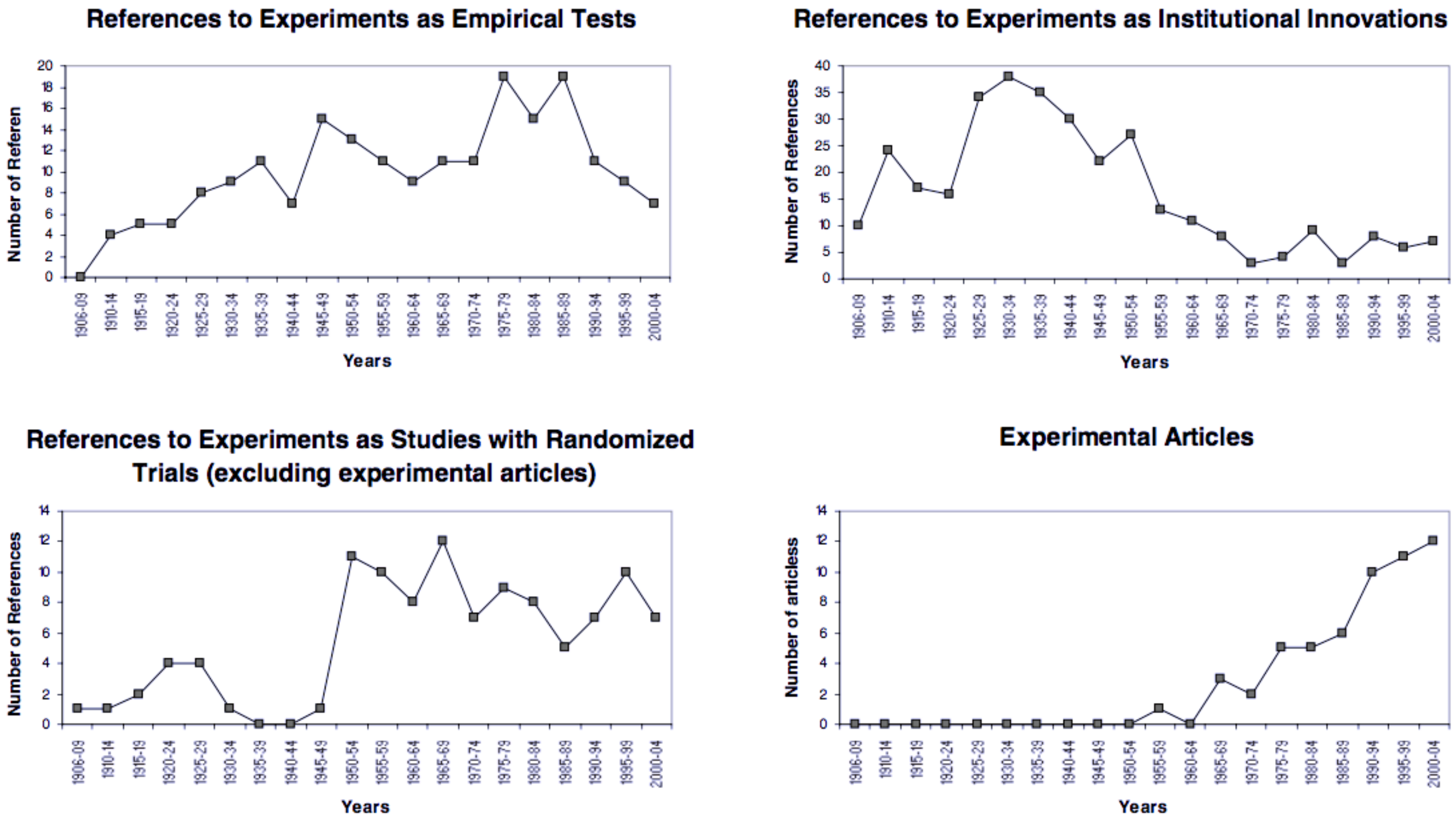
# Why experimental design?

**FIGURE 1. "Experimental" Trends in the *Review***



# Why experimental design?

**FIGURE 1. "Experimental" Trends in the *Review***




- Druckman et al. 2006

# Purposes of experiments

(Economist Alvin Roth)

# Purposes of experiments

## (Economist Alvin Roth)

1. **Searching for facts**

# Purposes of experiments

(Economist Alvin Roth)

1. **Searching for facts**
2. **Speaking to theory**

# Purposes of experiments

(Economist Alvin Roth)

1. **Searching for facts**
2. **Speaking to theory**
3. **Whispering in the ears of *princes***

# Purposes of experiments

## (Economist Alvin Roth)

# Purposes of experiments

(Economist Alvin Roth)

**1. Searching for facts**

“searching for facts,” where the goal is to “isolate the cause of some observed regularity, by varying details of the way the experiments were conducted. Such experiments are part of the dialog that experimenters carry on with one another.”

# Purposes of experiments

## (Economist Alvin Roth)

# Purposes of experiments

(Economist Alvin Roth)

**2. Speaking to theory**

where the goal is "to test the predictions [or the assumptions] of well articulated formal theories [or other types of theories]… Such experiments are intended to feed back into the theoretical literature––that is, they are part of a dialogue between experimenters and theorists."

# Purposes of experiments

## (Economist Alvin Roth)

# Purposes of experiments

(Economist Alvin Roth)

**3. whispering in the ears of princes**

which facilitates “the dialogue between experimenters and policymakers . . . [The] experimental environment is designed to resemble closely, in certain respects, the naturally occurring environment that is the focus of interest for the policy purposes at hand.”

# Observational vs. Experimental data

# Observational vs. Experimental data

In observational studies , researchers collect subject data and measure variables of interest without assigning treatments to the subjects.

# Observational vs. Experimental data

In observational studies , researchers collect subject data and measure variables of interest without assigning treatments to the subjects.

In an experiment investigators apply treatments to experimental units (subject) and then proceed to observe the effect of the treatments on the experimental units.

# Observational study

# Observational study

1. Find 100 women age 30 of which 50 watch TV every day while the other 50 do not.

# Observational study

1. Find 100 women age 30 of which 50 watch TV every day while the other 50 do not.
2. Measure political knowledge for each of the 100 women.

# Observational study

1. Find 100 women age 30 of which 50 watch TV every day while the other 50 do not.
2. Measure political knowledge for each of the 100 women.
3. Analyze, interpret, and draw conclusions from data.

# Experimental study

# Experimental study

1. Find 100 women age 30 who do not watch TV.

# Experimental study

1. Find 100 women age 30 who do not watch TV.
2. Measure political knowledge for each of the 100 women

# Experimental study

1. Find 100 women age 30 who do not watch TV.
2. Measure political knowledge for each of the 100 women
3. Randomly assign 50 of the 100 women to watch TV for 10 days and the other 50 remain not exposed to TV.

# Experimental study

1. Find 100 women age 30 who do not watch TV.
2. Measure political knowledge for each of the 100 women
3. Randomly assign 50 of the 100 women to watch TV for 10 days and the other 50 remain not exposed to TV.
4. Measure political knowledge for each of the 100 women again and analyze results.

# The Experimental Method

# The Experimental Method

1. What is an experiment?

# The Experimental Method

1. What is an experiment?

Purpose: The experimental design is to demonstrate causation, that A causes B

# The Experimental Method

1. What is an experiment?

Purpose: The experimental design is to demonstrate causation, that A causes B

- A -> B

# The Experimental Method

1. What is an experiment?

Purpose: The experimental design is to demonstrate causation, that A causes B

- A -> B

Requirements to demonstrate causality?

# The Experimental Method

1. What is an experiment?

Purpose: The experimental design is to demonstrate causation, that A causes B

- A -> B

Requirements to demonstrate causality?

- Correlation

# The Experimental Method

1. What is an experiment?

Purpose: The experimental design is to demonstrate causation, that A causes B

- A -> B

Requirements to demonstrate causality?

- Correlation
- Order. A must precede B.

# The Experimental Method

1. What is an experiment?

Purpose: The experimental design is to demonstrate causation, that A causes B

- A -> B

Requirements to demonstrate causality?

- Correlation
- Order. A must precede B.
- Control over other variables

# The Experimental Method

# The Experimental Method

2. How are experiments different from other types of research?

# The Experimental Method

2. How are experiments different from other types of research?

- Manipulated independent variable (treatment)

# The Experimental Method

2. How are experiments different from other types of research?

- Manipulated independent variable (treatment)
- Control of subject variables either by:

# The Experimental Method

2. How are experiments different from other types of research?

- Manipulated independent variable (treatment)
- Control of subject variables either by:
  - Random assignment of units of analysis to conditions of the independent variable, or

# The Experimental Method

2. How are experiments different from other types of research?

- Manipulated independent variable (treatment)
- Control of subject variables either by:
  - Random assignment of units of analysis to conditions of the independent variable, or
  - Assignment of each unit to all conditions, with controls on order of presentation

# The Experimental Method

# The Experimental Method

2. How are experiments different from other types of research? (continued)

# The Experimental Method

2. How are experiments different from other types of research? (continued)

- Control of other variables by holding them constant

# The Experimental Method

2. How are experiments different from other types of research? (continued)

- Control of other variables by holding them constant
- “In an airtight experiment, there is only one rival hypothesis: chance.”

# Strengths and Weaknesses of Experiments

# Strengths and Weaknesses of Experiments

| Strengths | Weaknesses |
|---|---|
| Control | Artificiality |
| Ability to demonstrate causality | Lack of external validity |

# On Internal Validity and External Validity

Internal validity addresses the question on whether the experimental treatments in fact make a difference in this specific experimental instance.

External validity regards the question of generalizability: to what populations, settings, treatment variables, and measurement variables can this effect be generalized. (Campbell and Stanley 1966)

# On External Validity of Experiments

# On External Validity of Experiments

"the conventional survey interview, though well equipped to assess variations among individuals, is poorly equipped to assess variation across situations."

- Sniderman et al. (1991: 265)

# On External Validity of Experiments

# On External Validity of Experiments

Unlike most controlled lab settings, researchers using survey experiments have limited ability introduce *contextual* variations.

- Druckman and Kam 2009

# Compare experiments with other data methods

# Compare experiments with other data methods

- Web data

# Compare experiments with other data methods

- Web data
- Observations

# Compare experiments with other data methods

- Web data
- Observations
- Survey

# Compare experiments with other data methods

- Web data
- Observations
- Survey
- Expert interviews

# Experimental design

# Experimental design

- Control

# Experimental design

- Control
- Sampling via randomization

# Experimental design

- Control
- Sampling via randomization
- Learn from repeated experiments

# Experimental design

- Control
- Sampling via randomization
- Learn from repeated experiments
  - Adaptive Clinical Design

# Experimental design

- Control
- Sampling via randomization
- Learn from repeated experiments
    - Adaptive Clinical Design
- Machine Learning

# Data quality

# Data quality

is a function of:

# Data quality

is a function of:

- Theory-driven

# Data quality

is a function of:

- Theory-driven
- Causality

# Data quality

is a function of:

- Theory-driven
- Causality
- Design

# Data quality

is a function of:

- Theory-driven
- Causality
- Design
- Control

# Data quality

is a function of:

- Theory-driven
- Causality
- Design
- Control
- Sampling

# Statistical Learning Process

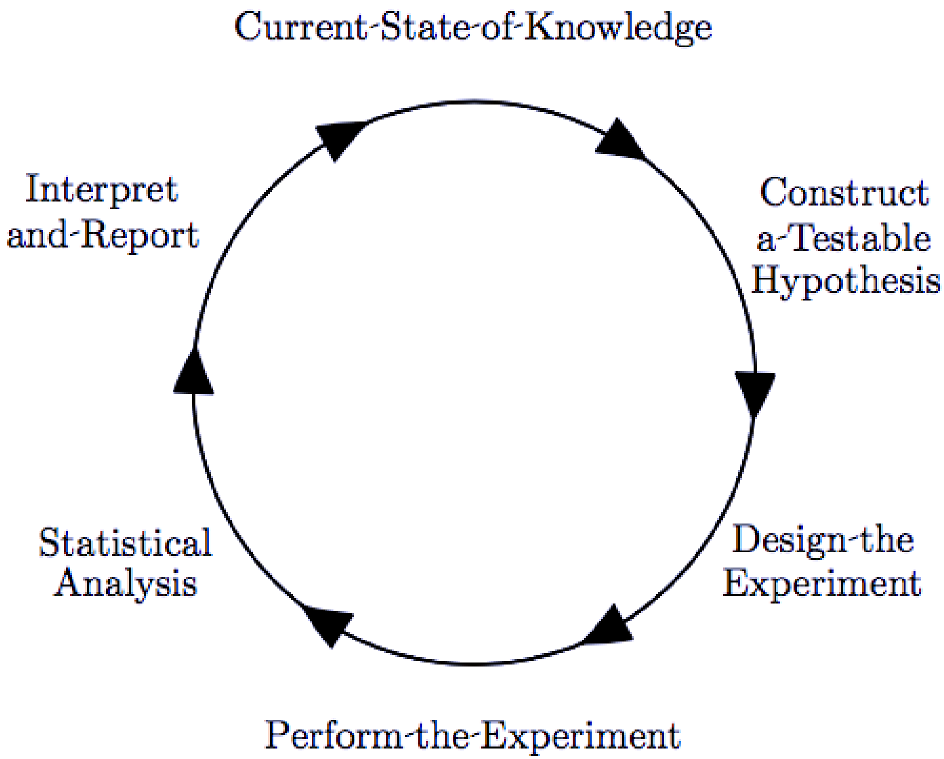



Figure 1.1: The circular flow of scientific learning

Source: Seltman, H.J., 2012. Experimental design and analysis. *Pittsburgh: Carnegie Mellon University*, *428*.

# Reference


Campbell, D.T. and Stanley, J.C., 1966. Experimental and quasi-experimental designs for research. *Handbook of research on teaching* (NL Gage, Ed.), pp.171-246.

Druckman, J.N. and Kam, C.D., 2009. Students as experimental participants: A defense of the 'narrow data base'.

Morton, R.B. and Williams, K.C., 2010. *Experimental political science and the study of causality: From nature to the lab.* Cambridge University Press.

Seltman, H.J., 2012. *Experimental design and analysis*. Pittsburgh: Carnegie Mellon University, 428.

# Data Methods: Qualitative Data

Karl Ho
School of Economic, Political and Policy Sciences
University of Texas at Dallas

# What is Qualitative research?

Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretive, material practices that turn or convert the world into a series of representations, including field notes, interviews, conversations, photographs, recordings, and memos to the self.

(Guest, Namey and Mitchell 2017)

# What is Qualitative research?

In other words, qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them.

(Denzin & Lincoln, 2005, p. 3)

# What is Qualitative Data?

# What is Qualitative Data?

Qualitative research involves any research that uses data that do not indicate ordinal values.

(Nkwi, Nyamongo, and Ryan 2001, p. 1)

# What is Qualitative Data?

Qualitative research involves any research that uses data that do not indicate ordinal values.

(Nkwi, Nyamongo, and Ryan 2001, p. 1)

In short, qualitative research involves collecting and/or working with text, images, or sounds.

# Qualitative Data Methods

# Qualitative Data Methods

1. Participant observation

# Qualitative Data Methods

1. Participant observation
2. In-depth interviews

# Qualitative Data Methods

1. Participant observation
2. In-depth interviews
3. Focus groups

# Qualitative Data Methods

1. Participant observation
2. In-depth interviews
3. Focus groups
4. Systematic elicitation

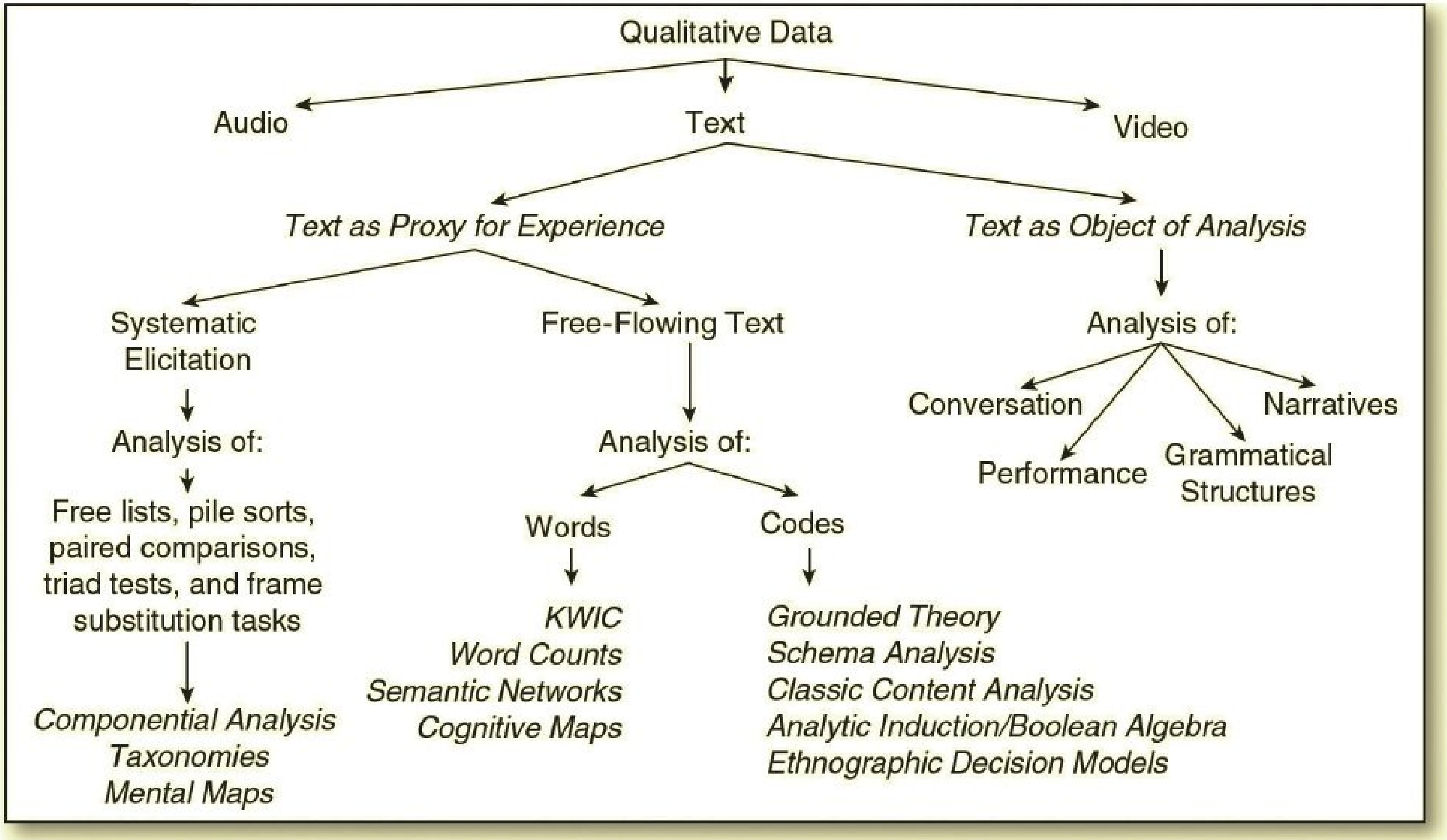
# Qualitative Data Methods

1. Participant observation
2. In-depth interviews
3. Focus groups
4. Systematic elicitation
5. Document analysis

# Qualitative Research and Data

# Qualitative Research and Data

**Figure 1.1 Typology of Qualitative Research**




*Source:* Ryan and Bernard (2000).

# Objectives of Qualitative Research

# Objectives of Qualitative Research

1. Identifying and exploring

# Objectives of Qualitative Research

1. **Identifying and exploring**
    - Building a list

# Objectives of Qualitative Research

1. Identifying and exploring
    - Building a list
2. Describing

# Objectives of Qualitative Research

1. Identifying and exploring
   - Building a list
2. Describing
   - Deep data

# Objectives of Qualitative Research

1. Identifying and exploring
   - Building a list
2. Describing
   - Deep data
3. Explaining

# Objectives of Qualitative Research

1. **Identifying and exploring**
   - Building a list
2. **Describing**
   - Deep data
3. **Explaining**
   - Addressing whys

# Qualitative Data Methods

# Qualitative Data Methods

1. Inductive approach

# Qualitative Data Methods

1. Inductive approach
2. How and why questions

# Qualitative Data Methods

1. Inductive approach
2. How and why questions
3. Open-end exploratory questions

# Qualitative Data Methods

1. Inductive approach
2. How and why questions
3. Open-end exploratory questions
4. Sequence:

# Qualitative Data Methods

1. Inductive approach
2. How and why questions
3. Open-end exploratory questions
4. Sequence:
    1. What do you think?

# Qualitative Data Methods

1. Inductive approach
2. How and why questions
3. Open-end exploratory questions
4. Sequence:
    1. What do you think?
    2. Why do you think so?

# Qualitative Data Methods

1. Inductive approach
2. How and why questions
3. Open-end exploratory questions
4. Sequence:
    1. What do you think?
    2. Why do you think so?
    3. How do you like or dislike_____?

# Qualitative Research: Human Experience

# Qualitative Research: Human Experience

1. Behaviors

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions
3. Knowledge

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions
3. Knowledge
4. Emotions and Values

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions
3. Knowledge
4. Emotions and Values
5. Culturally Shared Meaning

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions
3. Knowledge
4. Emotions and Values
5. Culturally Shared Meaning
6. Social Structures and Relationships

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions
3. Knowledge
4. Emotions and Values
5. Culturally Shared Meaning
6. Social Structures and Relationships
7. Processes and Systems

# Qualitative Research: Human Experience

1. Behaviors
2. Attitudes/Opinions/Perceptions
3. Knowledge
4. Emotions and Values
5. Culturally Shared Meaning
6. Social Structures and Relationships
7. Processes and Systems
8. Environmental Context

# Qualitative Research: Temporal dimension

# Qualitative Research: Temporal dimension

1. Single time point (cross-sectional)

# Qualitative Research: Temporal dimension

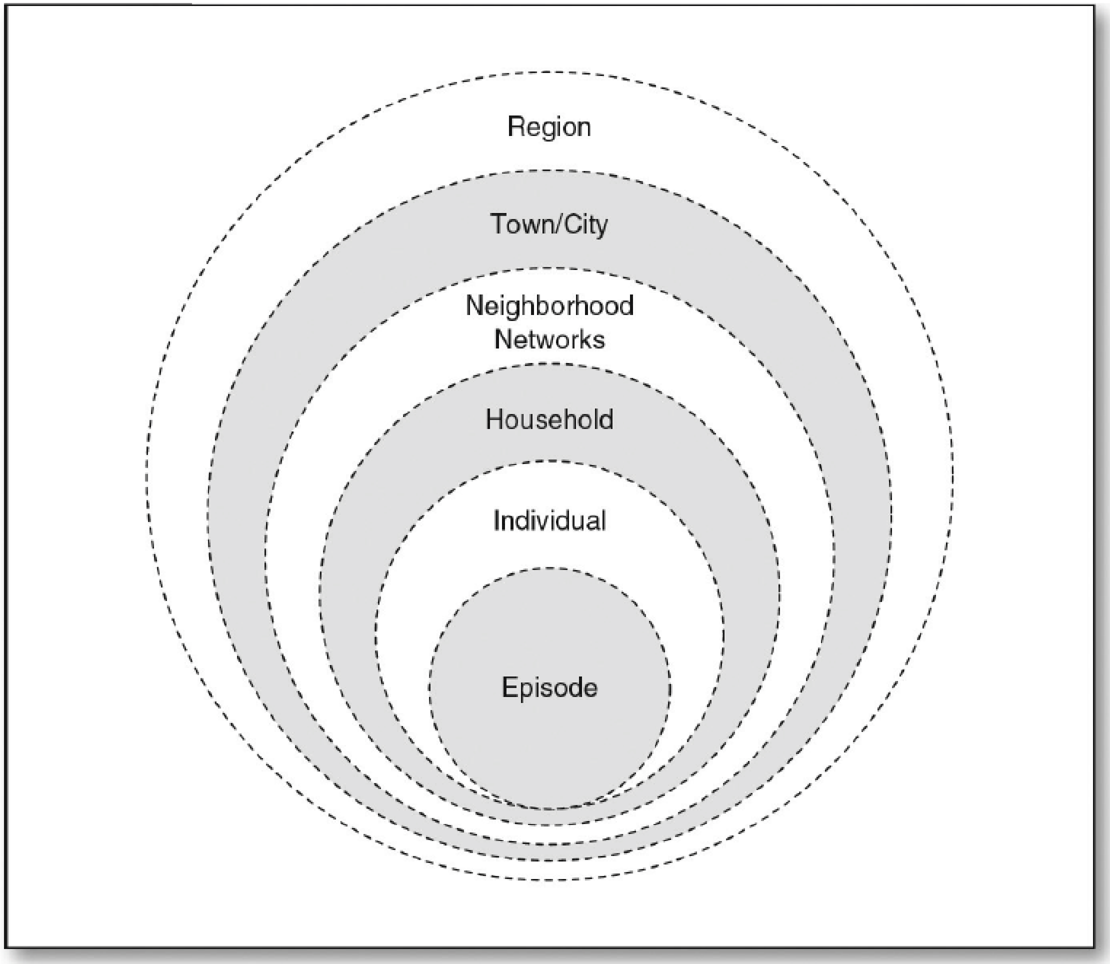
1. Single time point (cross-sectional)
2. Longitudinal study

# Qualitative Research: Temporal dimension

1. Single time point (cross-sectional)
2. Longitudinal study
3. Panel study (cohort study)

# Qualitative Research: Level and Unit of Analysis

**Figure 1.2 Levels of Analysis**




*Source:* Bernard and Ryan (2010, p. 129).

# Open interviews

# Open interviews

1. Causality

# Open interviews

1. Causality
2. Face validity

# Open interviews

1. Causality
2. Face validity
3. Reliability issue

# Open interviews

1. Causality
2. Face validity
3. Reliability issue
4. Future:

# Open interviews

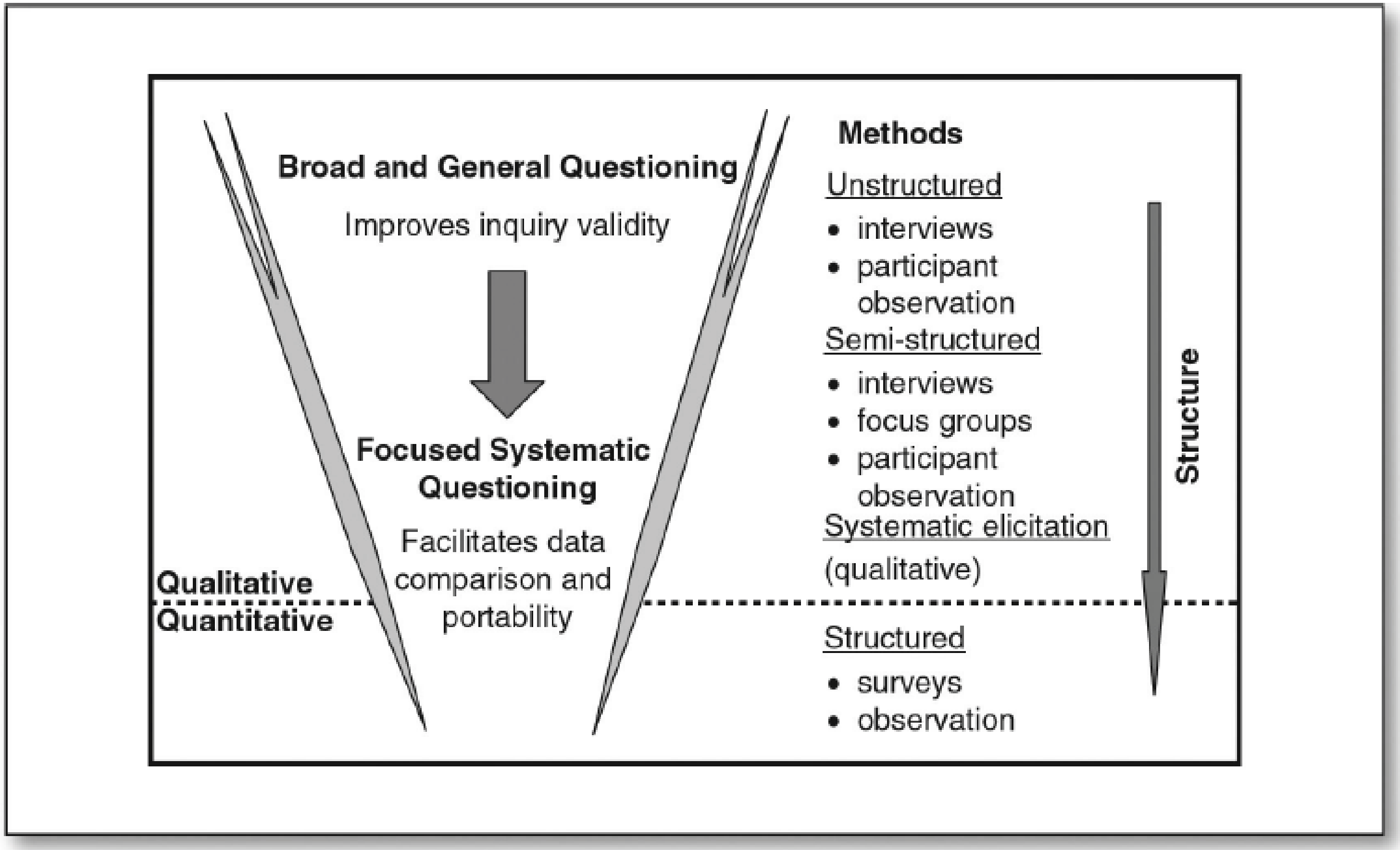
1. Causality
2. Face validity
3. Reliability issue
4. Future:
    1. AI guided internet survey

# Open interviews

1. Causality
2. Face validity
3. Reliability issue
4. Future:
    1. AI guided internet survey
    2. Accumulated data using machine-learning

# Qualitative Research and structure

Figure 1.3 The Research Process and Degree of Structure



# Qualitative Data: Sampling

# Qualitative Data: Sampling

1. What is the population?

# Qualitative Data: Sampling

1. **What is the population?**
    1. What is the representative sample?

# Qualitative Data: Sampling

1. **What is the population?**
    1. What is the representative sample?
2. **Approaches**

# Qualitative Data: Sampling

1. **What is the population?**
   1. What is the representative sample?
2. **Approaches**
3. **Sample size**

# Qualitative Data: Sampling Strategy

# Qualitative Data: Sampling Strategy

1. Estimate the size of the population of interest

# Qualitative Data: Sampling Strategy

1. Estimate the size of the population of interest
2. Control

# Qualitative Data: Sampling Strategy

1. **Estimate the size of the population of interest**
2. **Control**
   1. how much control you will have over your recruitment and sampling procedures

# Qualitative Data: Sampling Strategy

1. **Estimate the size of the population of interest**
2. **Control**
    1. how much control you will have over your recruitment and sampling procedures
    2. how certain you are about who, what, where you need to sample for your study.

# Qualitative Data: Sampling Strategy

1. **Estimate the size of the population of interest**
2. **Control**
    1. how much control you will have over your recruitment and sampling procedures
    2. how certain you are about who, what, where you need to sample for your study.
3. **Accumulative**

# Qualitative Data: Sampling Strategy

1. **Estimate the size of the population of interest**
2. **Control**
    1. how much control you will have over your recruitment and sampling procedures
    2. how certain you are about who, what, where you need to sample for your study.
3. **Accumulative**
4. **Adaptive**

# Future of Qualitative Research

# Future of Qualitative Research

1. Other data than text

# Future of Qualitative Research

1. **Other data than text**
    1. Complex data

# Future of Qualitative Research

1. **Other data than text**
    1. Complex data
    2. Audio, visual and video data

# Future of Qualitative Research

1. **Other data than text**
    1. Complex data
    2. Audio, visual and video data
    3. Machine Learning

# Future of Qualitative Research

1. **Other data than text**
    1. Complex data
    2. Audio, visual and video data
    3. Machine Learning
    4. Natural Language Processing (NLP)

# Mixed-methods Research

# Mixed-methods Research

1. Integration of qualitative and quantitative methods in a single study

# Mixed-methods Research

1. Integration of qualitative and quantitative methods in a single study
2. Strategy instead of a method

# Mixed-methods Research

1. Integration of qualitative and quantitative methods in a single study
2. Strategy instead of a method
3. Attend to structure of data

# Mixed-methods Research: Sampling

# Mixed-methods Research: Sampling

# Mixed-methods Research: Sampling

1. Temporal dimension

# Mixed-methods Research: Sampling

1. Temporal dimension
   1. Sequential

# Mixed-methods Research: Sampling

1. Temporal dimension
   1. Sequential
   2. Concurrent

# Mixed-methods Research: Sampling

1. Temporal dimension
   1. Sequential
   2. Concurrent
2. Relationships of Samples

# Mixed-methods Research: Sampling

1. **Temporal dimension**
   1. Sequential
   2. Concurrent
2. **Relationships of Samples**
   1. Identical

# Mixed-methods Research: Sampling

1. **Temporal dimension**
   1. Sequential
   2. Concurrent
2. **Relationships of Samples**
   1. Identical
   2. Parallel

# Mixed-methods Research: Sampling

1. **Temporal dimension**
   1. Sequential
   2. Concurrent
2. **Relationships of Samples**
   1. Identical
   2. Parallel
   3. Nested

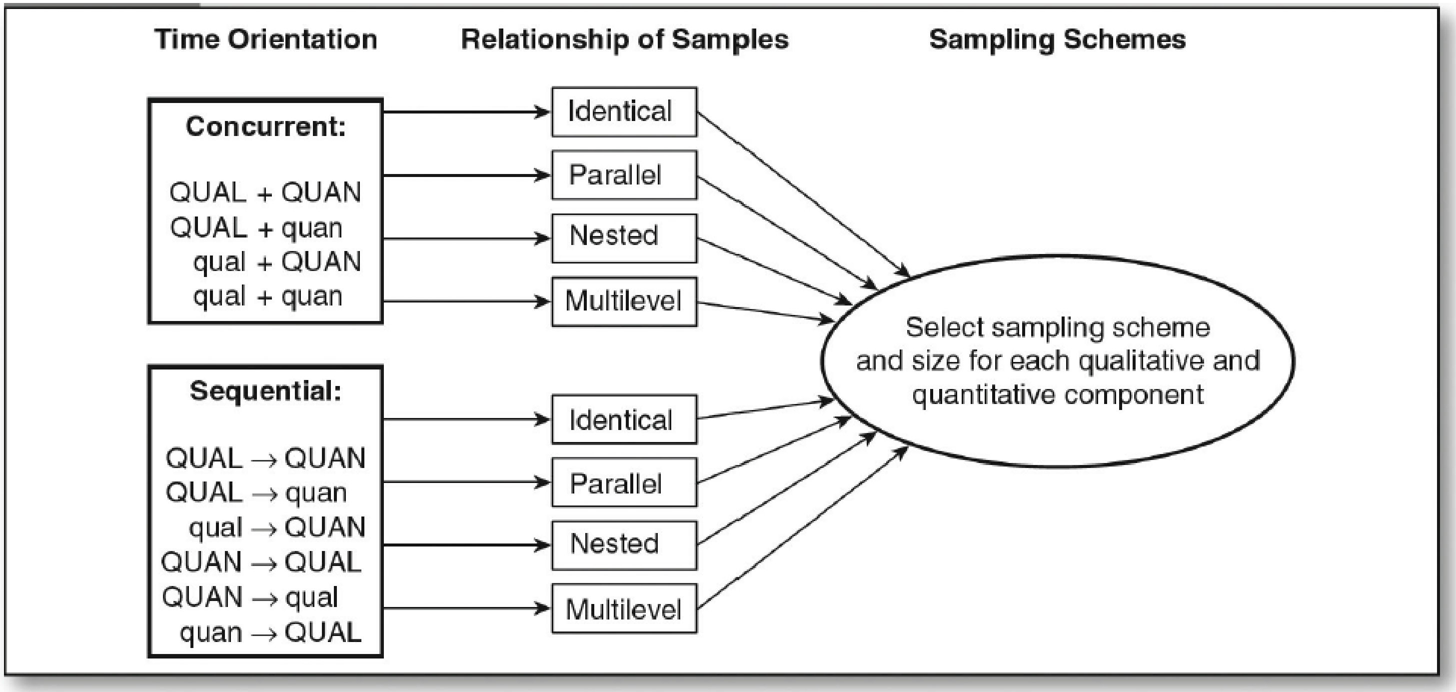
# Mixed-methods Research: Sampling

1. **Temporal dimension**
    1. Sequential
    2. Concurrent
2. **Relationships of Samples**
    1. Identical
    2. Parallel
    3. Nested
    4. Multilevel

# Mixed-methods Research

# Mixed-methods Research

**Figure 2.2 Mixed Methods Sampling Approaches**




*Source:*Onwuegbuzie and Collins (2007, p. 294).

# Managing Qualitative Data

# Managing Qualitative Data

Data management is “a designed structure for systematizing, categorizing, and filing materials to make them efficiently retrievable and duplicable”

- Schwandt, 1997, p. 61

# Data Methods:
# Focus Group Research

Karl Ho
School of Economic, Political and Policy Sciences
University of Texas at Dallas

# What is Focus Group?

# What is Focus Group?

A carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment.

- Krueger & Casey

# What is Focus Group?

# What is Focus Group?

Each group is conducted with six to eight people by a skilled interviewer. The discussions are relaxed, and often participants enjoy sharing their ideas and perceptions. Group members influence each other by responding to ideas and comments of others.

- Krueger & Casey

# FGR Data Collection

# FGR Data Collection

- Qualitative Data

# FGR Data Collection

- Qualitative Data
- Small Data Approach

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding
- Sampling

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding
- Sampling
  - Homogeneous

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding
- Sampling
  - Homogeneous
  - Snowball sampling

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding
- Sampling
  - Homogeneous
  - Snowball sampling
  - IRB (sample)

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding
- Sampling
  - Homogeneous
  - Snowball sampling
  - IRB (sample)
- Data type:

# FGR Data Collection

- Qualitative Data
- Small Data Approach
- Researcher is data collector
- Pre-planned Instrument
- Coding
- Sampling
  - Homogeneous
  - Snowball sampling
  - IRB (sample)
- Data type:
  - Connected

# FGR Interviews
## Characteristics

# FGR Interviews

## Characteristics

1. Participants

# FGR Interviews

## Characteristics

1. Participants
2. Environment

# FGR Interviews

## Characteristics

1. Participants
2. Environment
3. Moderator

# FGR Interviews

## Characteristics

1. Participants
2. Environment
3. Moderator
4. Data and Report

# FGR Interviews

## Characteristics

# FGR Interviews
## Characteristics

1. Participants
   - Recruitment (sampling)

# FGR Interviews

## Characteristics

1. Participants
   - Recruitment (sampling)
   - 5 to 10 people per group, 6-8 preferred

# FGR Interviews

## Characteristics

1. Participants
   - Recruitment (sampling)
   - 5 to 10 people per group, 6-8 preferred
   - Homogeneity

# FGR Interviews Characteristics

1. Participants
    - Recruitment (sampling)
    - 5 to 10 people per group, 6-8 preferred
    - Homogeneity
    - Repeated groups

# FGR Interviews

## Characteristics

2. Environment

# FGR Interviews

## Characteristics

2. Environment

- Comfortable

# FGR Interviews

## Characteristics

2. Environment

- Comfortable
- Circle seating

# FGR Interviews
## Characteristics

2. Environment
    - Comfortable
    - Circle seating
    - Recording

# FGR Interviews

## Characteristics

3. Moderator

# FGR Interviews
## Characteristics

3. Moderator

- Skillful in group discussions

# FGR Interviews

## Characteristics

3. Moderator

- Skillful in group discussions
- Uses pre-determined questions

# FGR Interviews

## Characteristics

3. Moderator

- Skillful in group discussions
- Uses pre-determined questions
- Establishes permissive environment

# FGR Interviews

## Characteristics

4. Data and Reporting

# FGR Interviews

## Characteristics

4. Data and Reporting

- Identify Big idea

# FGR Interviews
## Characteristics

4. Data and Reporting
   - Identify Big idea
   - Systematic analysis

# FGR Interviews

## Characteristics

4. Data and Reporting

- Identify Big idea
- Systematic analysis
- Verifiable procedures

# FGR Interviews
## Characteristics

4. Data and Reporting

- Identify Big idea
- Systematic analysis
- Verifiable procedures
- Next step of data

# FGR Interviews Characteristics

4. Data and Reporting

- Identify Big idea
- Systematic analysis
- Verifiable procedures
- Next step of data
- Appropriate reporting

# FGR Data Discussions

# FGR Data Discussions

- Preparing data for analysis and replication

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias
- Interaction data

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias
- Interaction data
- Data Connectedness

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias
- Interaction data
- Data Connectedness
- Disciplines:

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias
- Interaction data
- Data Connectedness
- Disciplines:
  - Medical: practitioners such as doctors, nurses, therapists and researchers

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias
- Interaction data
- Data Connectedness
- Disciplines:
  - Medical: practitioners such as doctors, nurses, therapists and researchers
  - Social work and Counseling

# FGR Data Discussions

- Preparing data for analysis and replication
- Moderator bias
- Interaction data
- Data Connectedness
- Disciplines:
  - Medical: practitioners such as doctors, nurses, therapists and researchers
  - Social work and Counseling
  - Anthropology

# FGR Data Discussions

# FGR Data Discussions

- Issues

# FGR Data Discussions

- Issues
  - Inference

# FGR Data Discussions

- Issues
  - Inference
    - Theoretical Generalization vs. Empirical Generalization (Sim 1998)

# FGR Data Discussions

- Issues
  - Inference
    - Theoretical Generalization vs. Empirical Generalization (Sim 1998)
  - Presence vs. absence of explanatory factors

# FGR Data Discussions

- Issues
  - Inference
    - Theoretical Generalization vs. Empirical Generalization (Sim 1998)
  - Presence vs. absence of explanatory factors
  - Factor structure

# Reference (for further reading)


Sim, J., 1998. Collecting and analysing qualitative data: issues raised by the focus group. *Journal of advanced nursing*, 28(2), pp.345-352.

Saldaña, J., 2015. *The coding manual for qualitative researchers*. Sage.

# Software

# Software

- Nvivo

# Software

- Nvivo
- ATLAS.ti

# Software

- Nvivo
- ATLAS.ti
- Nud.ist

# Software

- Nvivo
- ATLAS.ti
- Nud.ist
- Ethnograph

# Software

- Nvivo
- ATLAS.ti
- Nud.ist
- Ethnograph
- Pralis

# Software

- Nvivo
- ATLAS.ti
- Nud.ist
- Ethnograph
- Pralis
- CAQDAS

# What is R?

# What is R?

- The R statistical programming language is a free, open source package based on the S language developed by John Chambers.

# What is R?

- The R statistical programming language is a free, open source package based on the S language developed by John Chambers.
- Some history of R and S

# What is R?

- The R statistical programming language is a free, open source package based on the S language developed by John Chambers.
- Some history of R and S
- It is:
  - Large, probably one of the largest based on the user-written add-ons/procedures
  - Object-oriented
  - Interactive
  - Multiplatform: Windows, Mac, Linux

# What is R?

According to John Chambers (2009), six facets of R :

# What is R?

According to John Chambers (2009), six facets of R :

1. an interface to computational procedures of many kinds;

# What is R?

According to John Chambers (2009), six facets of R :

1. an interface to computational procedures of many kinds;
2. interactive, hands-on in real time;

# What is R?

According to John Chambers (2009), six facets of R :

1. an interface to computational procedures of many kinds;
2. interactive, hands-on in real time;
3. functional in its model of programming;

# What is R?

According to John Chambers (2009), six facets of R :

1. an interface to computational procedures of many kinds;
2. interactive, hands-on in real time;
3. functional in its model of programming;
4. object-oriented, “everything is an object”;

# What is R?

According to John Chambers (2009), six facets of R :

1. an interface to computational procedures of many kinds;
2. interactive, hands-on in real time;
3. functional in its model of programming;
4. object-oriented, “everything is an object”;
5. modular, built from standardized pieces; and,

# What is R?

According to John Chambers (2009), six facets of R :

1. an interface to computational procedures of many kinds;
2. interactive, hands-on in real time;
3. functional in its model of programming;
4. object-oriented, “everything is an object”;
5. modular, built from standardized pieces; and,
6. collaborative, a world-wide, open-source effort.

# Why R?

# Why R?

- A programming platform environment

# Why R?

- A programming platform environment
- Allow development of software/packages by users

# Why R?

- A programming platform environment
- Allow development of software/packages by users
- Currently, the CRAN package repository features 12,108 available packages (as of 1/31/2018).

# Why R?

- A programming platform environment
- Allow development of software/packages by users
- Currently, the CRAN package repository features 12,108 available packages (as of 1/31/2018).
- Graphics!!!

# Why R?

- A programming platform environment
- Allow development of software/packages by users
- Currently, the CRAN package repository features 12,108 available packages (as of 1/31/2018).
- Graphics!!!
- Comparing R with other software?

# Why R?

| Features | Stata | SPSS | SAS | R |
|---|---|---|---|---|
| Learning curve | Steep/gradual | Gradual/flat | Pretty steep | Pretty steep |
| User interface | Programming/point-and-click | Mostly point-and-click | Programming | Programming |
| Data manipulation | Very strong | Moderate | Very strong | Very strong |
| Data analysis | Powerful | Powerful | Powerful/versatile | Powerful/versatile |
| Graphics | Very good | Very good | Good | Excellent |
| Cost | Affordable (perpetual licenses, renew only when upgrade) | Expensive (but not need to renew until upgrade, long term licenses) | Expensive (yearly renewal) | Open source |

Source: Oscar Torres-Reyna. 2010. *Getting Started in R~Stata Notes on Exploring Data,* (http://dss.Princeton.edu/training)

| Features | Stata | SPSS | SAS | R |
| --- | --- | --- | --- | --- |
| Learning curve | Steep/gradual | Gradual/flat | Pretty steep | Pretty steep |
| User interface | Programming/point-and-click | Mostly point-and-click | Programming | Programming |
| Data manipulation | Very strong | Moderate | Very strong | Very strong |
| Data analysis | Powerful | Powerful | Powerful/versatile | Powerful/versatile |
| Graphics | Very good | Very good | Good | Excellent |
| Cost | Affordable (perpetual licenses, renew only when upgrade) | Expensive (but not need to renew until upgrade, long term licenses) | Expensive (yearly renewal) | Open source |

# Why R?

| Features | Stata | SPSS | SAS | R |
|---|---|---|---|---|
| Learning curve | Steep/gradual | Gradual/flat | Pretty steep | Pretty steep |
| User interface | Programming/point-and-click | Mostly point-and-click | Programming | Programming |
| Data manipulation | Very strong | Moderate | Very strong | Very strong |
| Data analysis | Powerful | Powerful | Powerful/versatile | Powerful/versatile |
| Graphics | Very good | Very good | Good | Excellent |
| Cost | Affordable (perpetual licenses, renew only when upgrade) | Expensive (but not need to renew until upgrade, long term licenses) | Expensive (yearly renewal) | Open source |

Source: Oscar Torres-Reyna. 2010. *Getting Started in R~Stata Notes on Exploring Data,* (http://dss.Princeton.edu/training)

# Getting the software

# Getting the software

- R home: https://www.r-project.org

# Getting the software

- R home: https://www.r-project.org
- Downloads: https://cran.r-project.org

# Getting the software

- R home: https://www.r-project.org
- Downloads: https://cran.r-project.org
- Select Download R for Windows or Mac or Linux

# Getting the software

- R home: https://www.r-project.org
- Downloads: https://cran.r-project.org
- Select Download R for Windows or Mac or Linux
- R 3.4.3 "Kite-Eating Tree" released on 2017/11/30

# Getting the software

- R home: https://www.r-project.org
- Downloads: https://cran.r-project.org
- Select Download R for Windows or Mac or Linux
- R 3.4.3 "Kite-Eating Tree" released on 2017/11/30
- Caveat:
  - Frequent updates/upgrades
  - Packages/library

# Recommended R resources

# Recommended R resources

- The R Journal (http://journal.r-project.org/)

# Recommended R resources

- The R Journal (http://journal.r-project.org/)
- Introduction to R by W. N. Venables, D. M. Smith and the R Core Team (http://cran.r-project.org/doc/manuals/R-intro.pdf)

# Recommended R resources

- The R Journal (http://journal.r-project.org/)
- Introduction to R by W. N. Venables, D. M. Smith and the R Core Team (http://cran.r-project.org/doc/manuals/R-intro.pdf)
- Introduction to R Seminar at UCLA (http://www.ats.ucla.edu/stat/r/seminars/intro.htm)

# RStudio

RStudio is a user interface for the statistical programming software R.

# RStudio

RStudio is a user interface for the statistical programming software R.

- Object-based environment

# RStudio

RStudio is a user interface for the statistical programming software R.

- Object-based environment
- Window system

# RStudio

RStudio is a user interface for the statistical programming software R.

- Object-based environment
- Window system
- Point and click operations

# RStudio

RStudio is a user interface for the statistical programming software R.

- Object-based environment
- Window system
- Point and click operations
- Coding recommended

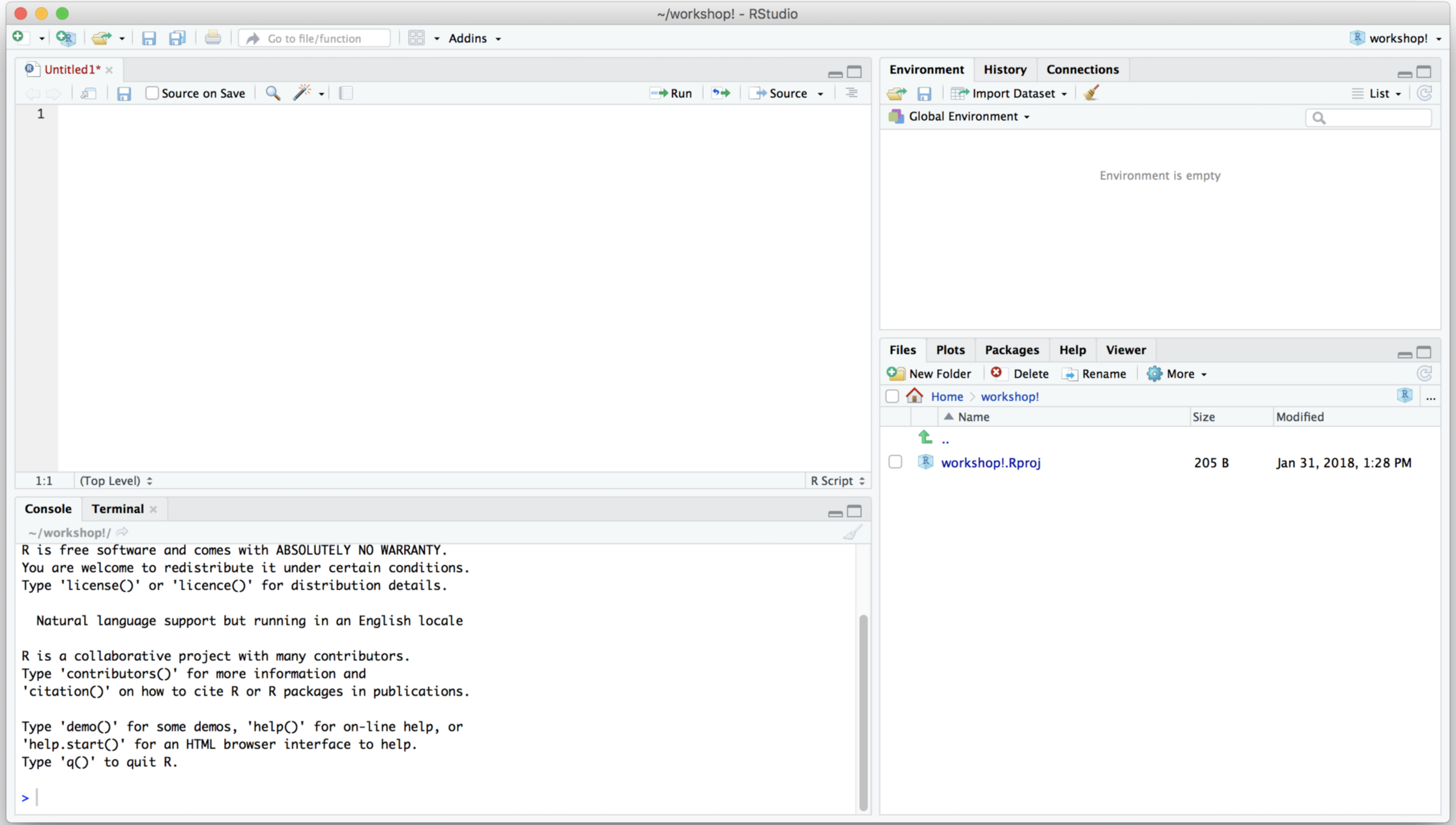
# RStudio

RStudio is a user interface for the statistical programming software R.

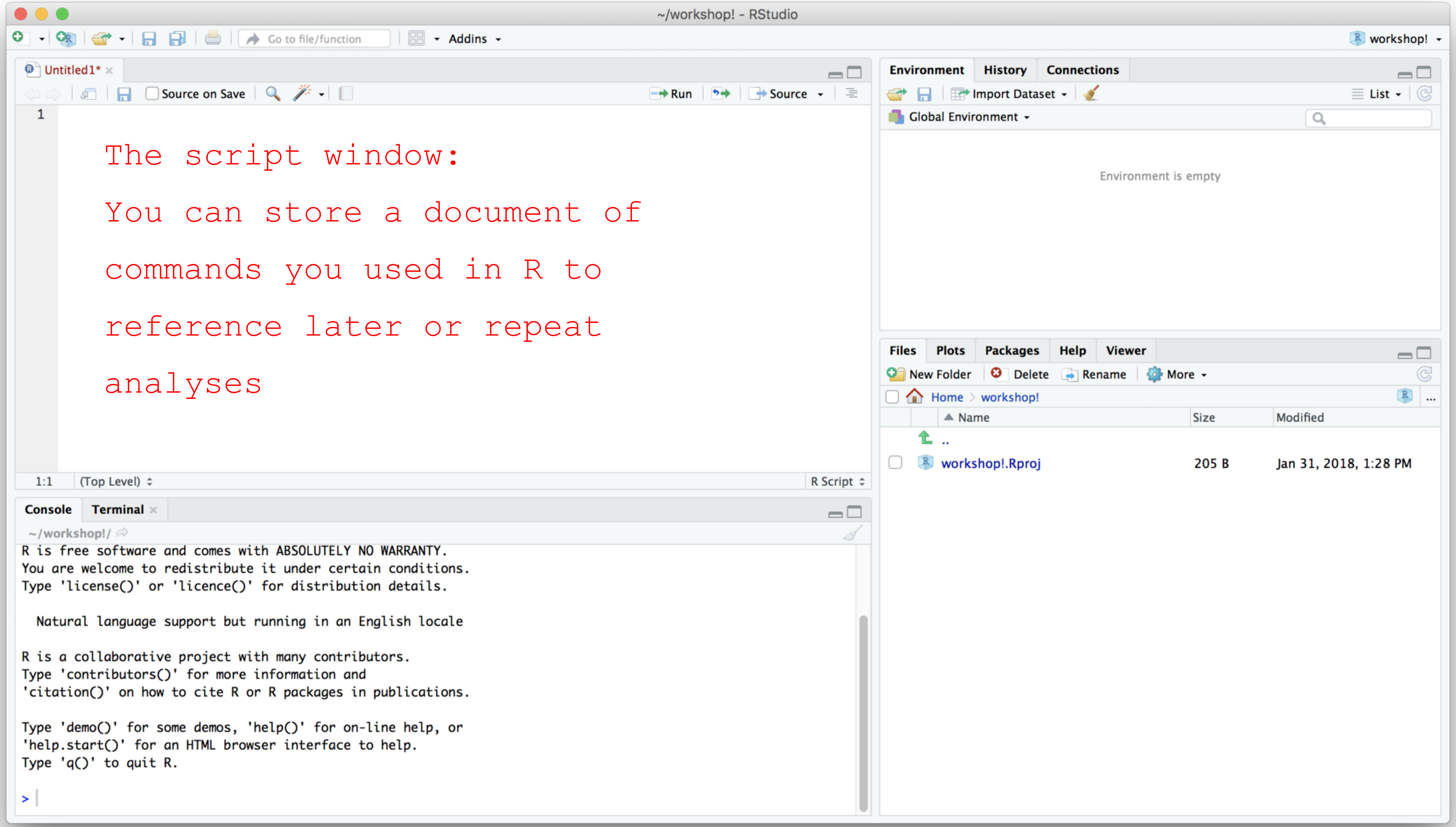
- Object-based environment
- Window system
- Point and click operations
- Coding recommended
- Expansions and development

# RStudio

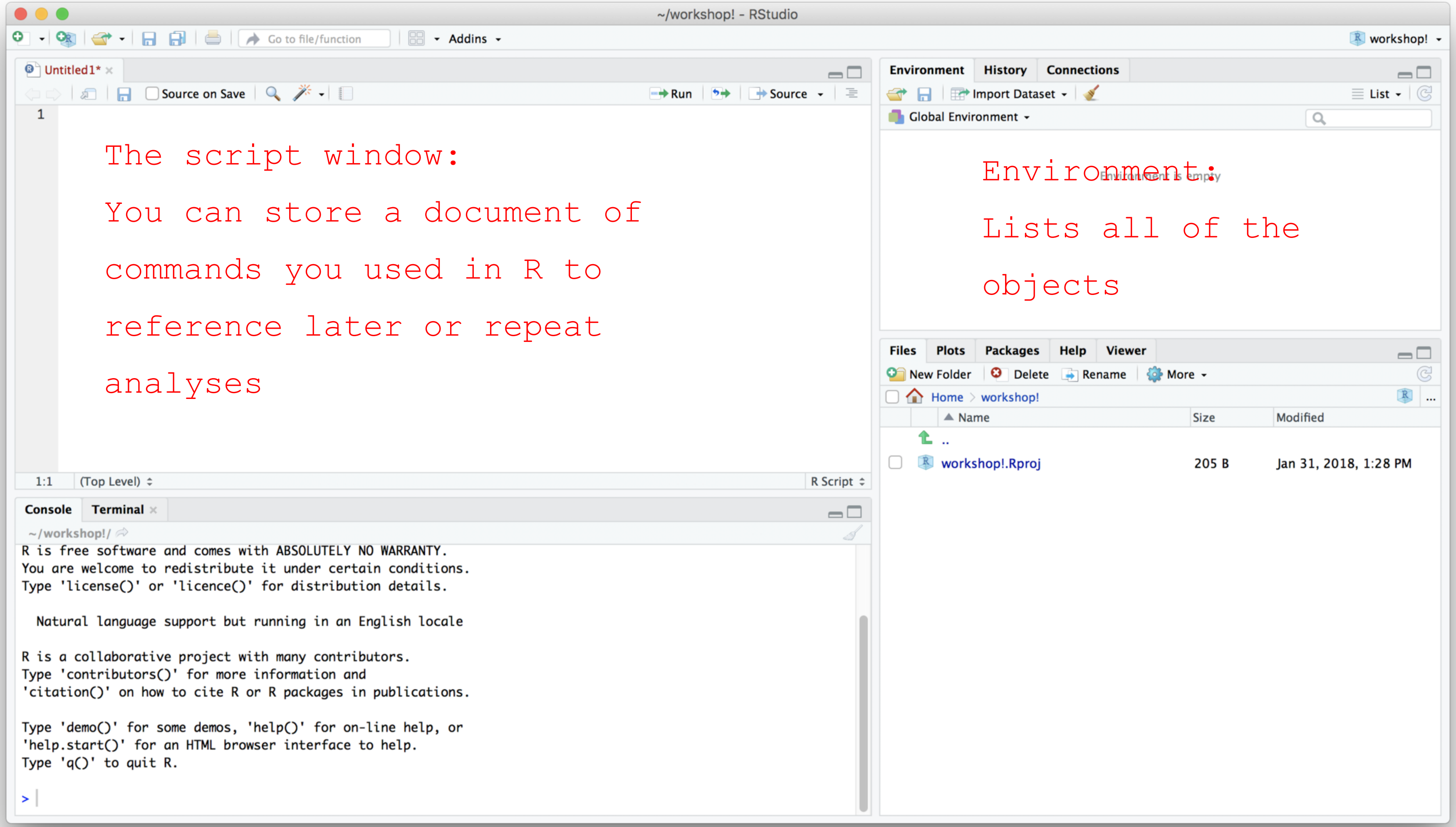
# RStudio


~/workshop! - RStudio
Go to file/function
Addins
workshop!
Untitled1*
Source on Save
Run
Source
1
1:1
(Top Level)
R Script
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
workshop!
Name
Size
Modified
..
workshop!.Rproj
205 B
Jan 31, 2018, 1:28 PM
Console
Terminal
~/workshop!/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>

# RStudio

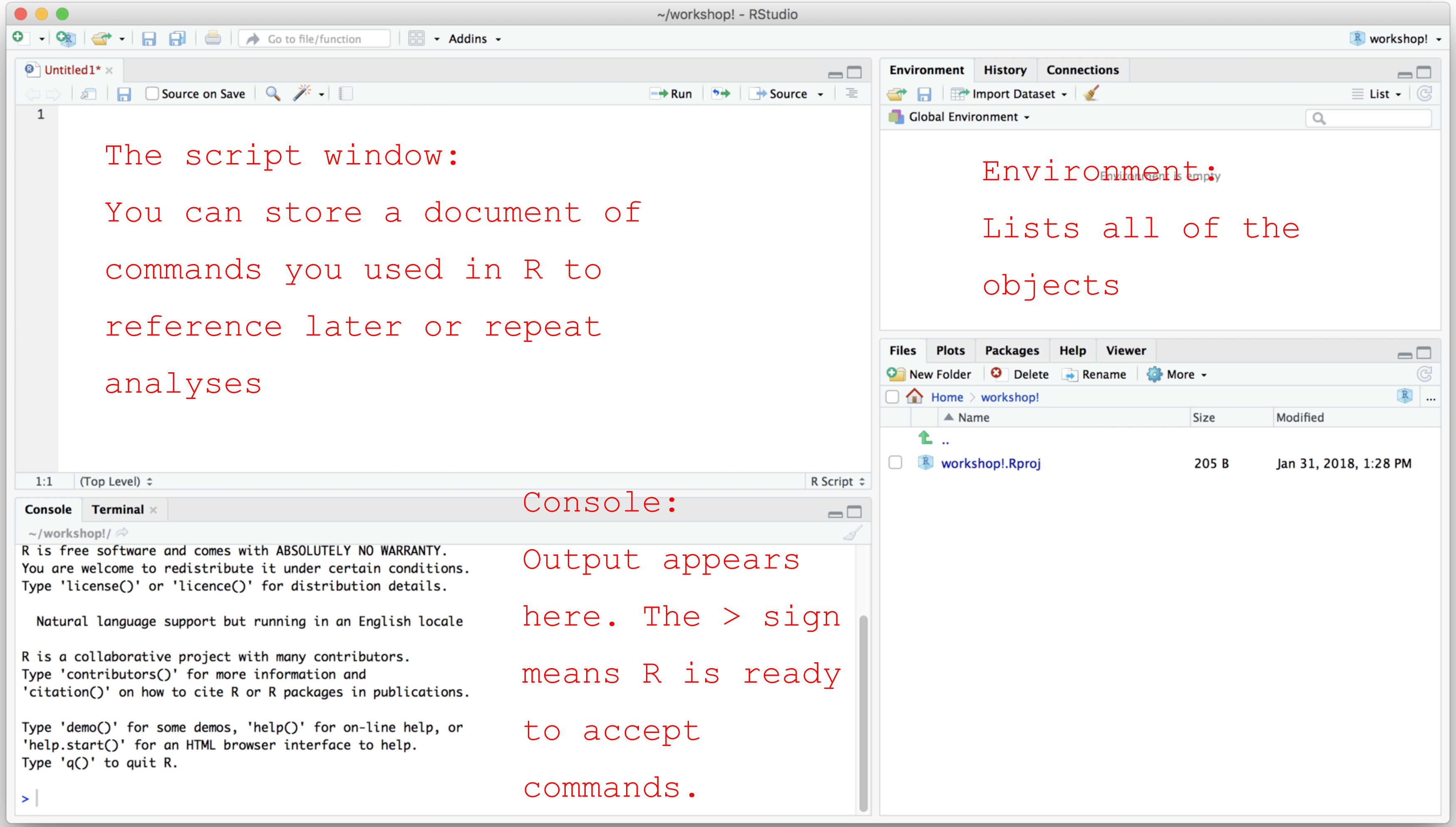

~/workshop! - RStudio
Go to file/function
Addins
workshop!
Untitled1*
Source on Save
Run
Source
1
The script window:
You can store a document of
commands you used in R to
reference later or repeat
analyses
1:1
(Top Level)
R Script
Console
Terminal
~/workshop!/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
workshop!
Name
Size
Modified
..
workshop!.Rproj
205 B
Jan 31, 2018, 1:28 PM

# RStudio

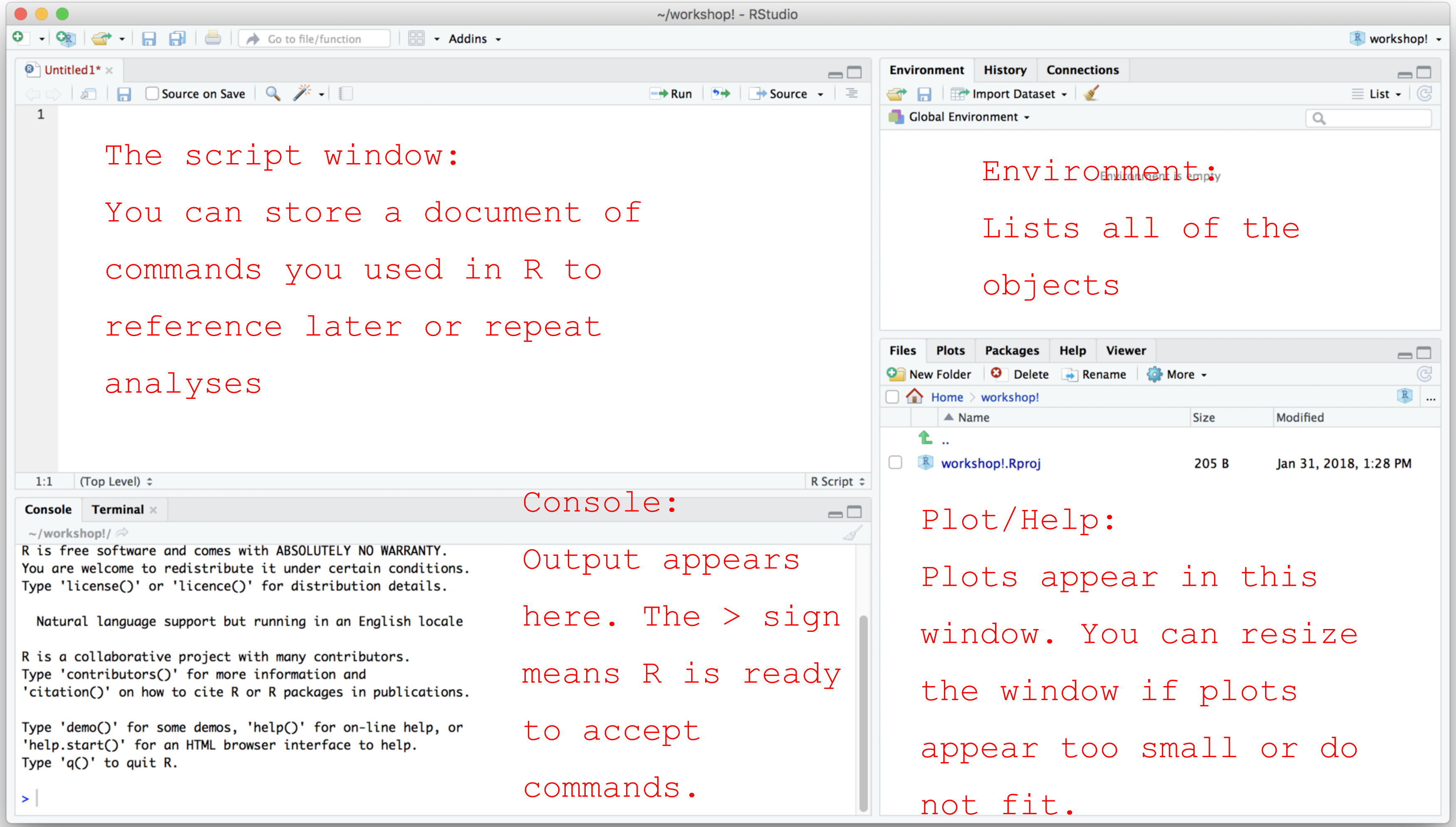


~/workshop! - RStudio
Go to file/function
Addins
workshop!
Untitled1*
Source on Save
Run
Source
1
The script window:
You can store a document of commands you used in R to reference later or repeat analyses
1:1
(Top Level)
R Script
Environment
History
Connections
Import Dataset
List
Global Environment
Environment:
Lists all of the objects
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
workshop!
Name
Size
Modified
workshop!.Rproj
205 B
Jan 31, 2018, 1:28 PM
Console
Terminal
~/workshop!/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>

# RStudio


~/workshop! - RStudio
Go to file/function
Addins
workshop!
Untitled1*
Source on Save
Run
Source
The script window:
You can store a document of commands you used in R to reference later or repeat analyses
1:1
(Top Level)
R Script
Console
Terminal
~/workshop!/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Console:
Output appears here. The > sign means R is ready to accept commands.
Environment
History
Connections
Import Dataset
List
Global Environment
Environment:
Lists all of the objects
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
workshop!
Name
Size
Modified
..
workshop!.Rproj
205 B
Jan 31, 2018, 1:28 PM

# RStudio


~/workshop! - RStudio
Go to file/function
Addins
workshop!
Untitled1*
Source on Save
Run
Source
1
The script window:
You can store a document of commands you used in R to reference later or repeat analyses
1:1
(Top Level)
R Script
Environment
History
Connections
Import Dataset
List
Global Environment
Environment:
Lists all of the objects
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
workshop!
Name
Size
Modified
..
workshop!.Rproj
205 B
Jan 31, 2018, 1:28 PM
Console
Terminal
~/workshop!/
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Console:
Output appears here. The > sign means R is ready to accept commands.
Plot/Help:
Plots appear in this window. You can resize the window if plots appear too small or do not fit.

# Getting started

# Getting started

- Basic assignment and operations.

# Getting started

- Basic assignment and operations.
- Arithmetic Operations:
  - +, -, *, /, ^ are the standard arithmetic operators.

# Getting started

- Basic assignment and operations.
- Arithmetic Operations:
  - +, -, *, /, ^ are the standard arithmetic operators.
- Assignment
  - To assign a value to a variable use “<-” or “=”

# R Programming Basics

# R Programming Basics

- R code can be entered into the command line directly or saved to a script, which can be run as a script

# R Programming Basics

- R code can be entered into the command line directly or saved to a script, which can be run as a script
- Commands are separated either by a ; or by a newline.

# R Programming Basics

- R code can be entered into the command line directly or saved to a script, which can be run as a script
- Commands are separated either by a ; or by a newline.
- R is case sensitive.

# R Programming Basics

- R code can be entered into the command line directly or saved to a script, which can be run as a script
- Commands are separated either by a ; or by a newline.
- R is case sensitive.
- The # character at the beginning of a line signifies a comment, which is not executed.

# R Programming Basics

- R code can be entered into the command line directly or saved to a script, which can be run as a script
- Commands are separated either by a ; or by a newline.
- R is case sensitive.
- The # character at the beginning of a line signifies a comment, which is not executed.
- Help can be accessed by preceding the name of the function with ? (e.g. ?plot)

# Importing data

# Importing data

- Can import from SPSS, Stata and text data file
  Use a package called foreign:
  First, install.packages(“foreign”), then you can use following codes to import data:

# Importing data

- Can import from SPSS, Stata and text data file
  Use a package called foreign:
  First, install.packages(“foreign”), then you can use following codes to import data:

```
mydata <-
read.csv(“path”,sep=“,”,header=TRUE)
mydata.spss <-
read.spss(“path”,sep=“,”,header=TRUE)
mydata.dta <-
read.dta(“path”,sep=“,”,header=TRUE)
```

# Importing data

# Importing data

- R is absolutely case-sensitive

# Importing data

Note:

- R is absolutely case-sensitive
- R uses extra backslashes to recognize path

# Importing data

Note:

- R is absolutely case-sensitive
- R uses extra backslashes to recognize path
- Read data directly from Github:

# Importing data

Note:

- R is absolutely case-sensitive
- R uses extra backslashes to recognize path
- Read data directly from Github:

```
happy=read.csv("https://raw.githu
busercontent.com/kho7/SDS/master/
Programming%20Workshops/R/happy.c
sv")
```

# Accessing variables

# Accessing variables

## To select a column use:

```
mydata$column
```

## For example:

# Accessing variables

## To select a column use:

`mydata$column`

## For example:

```
> happy$country
 [1] Afghanistan            Albania                  Algeria          Angola
 [5] Argentina              Armenia                  Australia        Austria
 [9] Azerbaijan             Bahrain                  Bangladesh       Belarus
[13] Belgium                Belize                   Benin            Bolivia
[17] Bosnia and Herzegovina Botswana                 Brazil           Bulgaria
[21] Burkina Faso           Burundi                  Cambodia         Cameroon
[25] Canada                 Central African Republic Chad             Chile
[29] China                  Colombia                 Comoros          Congo
[33] Congo, Dem. Rep. of the Costa Rica              Cote d'Ivoire    Croatia
[37] Cuba                   Cyprus                   Czech Republic   Denmark
[41] Djibouti               Dominican Republic       Ecuador          Egypt
[45] El Salvador            Estonia                  Ethiopia         Finland
[49] France                 Georgia                  Germany          Ghana
[53] Greece                 Guatemala                Guinea           Guyana
```

# Manipulating variables

# Manipulating variables

Recoding variables

For example:

# Manipulating variables

Recoding variables

For example:

```
mydata$Age.rec<-recode(mydata$Age,
"18:19='18to19';
20:29='20to29';30:39='30to39'")
```