Intro to Social Science Data Analysis

Lecture 1:
Introduction to the Course & R

Christopher Gandrud

September 3, 2012

Christopher Gandrud

- **Email:** gandrud@yonsei.ac.kr

- **Office Hours:** 15:00-17:00 Wednesday (정208)
- **Open Door:** You can always come to my office or email me.

Christopher Gandrud

- **Email:** gandrud@yonsei.ac.kr

- **Office Hours:** 15:00-17:00 Wednesday (정208)
- **Open Door:** You can always come to my office or email me.

Christopher Gandrud

- **Email:** gandrud@yonsei.ac.kr

- **Office Hours:** 15:00-17:00 Wednesday (정208)
- **Open Door:** You can always come to my office or email me.

What should you call me?

Please call me Chris or Christopher

In this course we respect **knowledge** & **evidence**, not titles.

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Why take this course?

- Learn how to take **raw data**, **explore it**, and **present what you find**.
- This course **is hands on** and **practical**. (If you want a mathmatical introduction to statistics, I recommend taking a coure in the Statistics department.)
- The skills taught in this course are important for many **real-world** situations:

1. Academics
2. Government
3. Buisiness
4. Journalism

Specifics

- We will learn how do these things with the **R** statistical language in the program RStudio.
  - R is difficult when you start to use it, but it is very **powerfull** and being able to use it is a very **marketable skill**.
- Finally, the course is **not** about memorisation.
- It is about developing tools to **solve new and unexpected problems**.

Specifics

- We will learn how do these things with the **R** statistical language in the program RStudio.
  - R is difficult when you start to use it, but it is very **powerfull** and being able to use it is a very **marketable skill**.
- Finally, the course is **not** about memorisation.
- It is about developing tools to **solve new and unexpected problems**.

Specifics

- We will learn how do these things with the **R** statistical language in the program RStudio.
  - R is difficult when you start to use it, but it is very **powerfull** and being able to use it is a very **marketable skill**.
- Finally, the course is **not** about memorisation.
- It is about developing tools to **solve new and unexpected problems**.

Specifics

- We will learn how do these things with the **R** statistical language in the program RStudio.
  - R is difficult when you start to use it, but it is very **powerfull** and being able to use it is a very **marketable skill**.
- Finally, the course is **not** about memorisation.
- It is about developing tools to **solve new and unexpected problems**.

Course Aims (3)

This course is useful.

This course is intended for beginners.

- You should have good **basic computer skills** (Have used Microsoft Excel, for example.)
- You need to be **curious**: Why did that happen? How can I solve this problem?
- You need to be **patient**: Can't give up why you don't succeed the first time.

This course is intended for beginners.

- You should have good **basic computer skills** (Have used Microsoft Excel, for example.)
- You need to be **curious**: Why did that happen? How can I solve this problem?
- You need to be **patient**: Can't give up why you don't succeed the first time.

This course is intended for beginners.

- You should have good **basic computer skills** (Have used Microsoft Excel, for example.)
- You need to be **curious**: Why did that happen? How can I solve this problem?
- You need to be **patient**: Can't give up why you don't succeed the first time.

Part I: Introduction to Data Gathering & Management

- R Basics: installing, objects, assignment, functions
- Data structures
- Gathering data
- Replication!

Part I: Introduction to Data Gathering & Management

- R Basics: installing, objects, assignment, functions
- Data structures
- Gathering data
- Replication!

Part I: Introduction to Data Gathering & Management

- R Basics: installing, objects, assignment, functions
- Data structures
- Gathering data
- Replication!

Part I: Introduction to Data Gathering & Management

- R Basics: installing, objects, assignment, functions
- Data structures
- Gathering data
- Replication!

Part II: Basic Data Analysis & Visualisation

- Descriptive statistics
- Data visualisation
- Overview of statistical inference
- Statistical inference with large samples

Part II: Basic Data Analysis & Visualisation

- Descriptive statistics
- Data visualisation
- Overview of statistical inference
- Statistical inference with large samples

Part II: Basic Data Analysis & Visualisation

- Descriptive statistics
- Data visualisation
- Overview of statistical inference
- Statistical inference with large samples

Part II: Basic Data Analysis & Visualisation

- Descriptive statistics
- Data visualisation
- Overview of statistical inference
- Statistical inference with large samples

Part III: Introduction to Linear Regression

- Simple linear regression
- Multiple linear regression
- Presenting regression results

Part III: Introduction to Linear Regression

- Simple linear regression
- Multiple linear regression
- Presenting regression results

Part III: Introduction to Linear Regression

- Simple linear regression
- Multiple linear regression
- Presenting regression results

Part IV: Introduction to Linear Regression

Research Projects:

Use all of these skills.

Course Materials (1)

Blog: `http://yonsei-data-analysis.tumblr.com/`

Password Protected: YonseiData

Syllabus: `http://bit.ly/QwE4UM`

Reading

The main **text** is: *OpenIntro Statistics (first edition)*
It is **free** and can be downloaded here:
`http://www.openintro.org/stat/downloads.php`.

You might also want to get: Crawley, Michael J. 2005. *Statistics: An Introduction Using R*. Chichester: John Wiley Sons. Ltd.

Course Materials (3)

However, the course is more about **doing** than **consuming**.
So the focus is on **completing tasks**, not reading.

To help you complete tasks, we are building a course **Wiki**:
`http://bit.ly/NkdgfW`.

Assessment (1)

- 10% Class Attendance and Participation
- 40% 5 Short Assignments: Due weeks 3, 5, 7, 9, 11
- 50% Pair Research project (paper and presentation): Due Week 16

Assessment (1)

- 10% Class Attendance and Participation
- 40% 5 Short Assignments: Due weeks 3, 5, 7, 9, 11
- 50% Pair Research project (paper and presentation): Due Week 16

Assessment (1)

- 10% Class Attendance and Participation
- 40% 5 Short Assignments: Due weeks 3, 5, 7, 9, 11
- 50% Pair Research project (paper and presentation): Due Week 16

Attendance and Participation

You must **attend** all lectures and seminars.

You must **participate** in class discussions and activities.

- Each $1 + N$ absence $= -5$ participation points.

Attendance and Participation

You must **attend** all lectures and seminars.

You must **participate** in class discussions and activities.

- Each $1 + N$ absence $= -5$ participation points.

Assessment (3)

More details will be given in future classes about the Short Assignments & Research Project

Now

- 1. Look at the course blog (`http://yonsei-data-analysis.tumblr.com/`)

  Please post!

- 2. Install R if you want to use your own computer. (RECOMMENDED)

  You can find instructions on the course wiki page: `http://bit.ly/PBjDdw`

- 3. Open and play around with RStudio

Now

- 1. Look at the course blog (`http://yonsei-data-analysis.tumblr.com/`)

  Please post!

- 2. Install R if you want to use your own computer. (RECOMMENDED)

  You can find instructions on the course wiki page: `http://bit.ly/PBjDdw`

- 3. Open and play around with RStudio

Now

- 1. Look at the course blog (`http://yonsei-data-analysis.tumblr.com/`)

  Please post!

- 2. Install R if you want to use your own computer. (RECOMMENDED)

  You can find instructions on the course wiki page: `http://bit.ly/PBjDdw`

- 3. Open and play around with RStudio