

STAT3622: Data Visualization

Dr. Aijun Zhang

1 September 201



Department of 統計及精算學系
Statistics & Actuarial Science

Outline

- 1 Course Outline
- 2 Tentative Schedule
- 3 RStudio and Markdown

Administrative Details

Instructor:	Dr. Aijun Zhang Office: RR224 Phone: 3917 1984 Email: ajzhang@hku.hk
TA:	Mr. Jia You (RR116) Email: u3005315@hku.hk
Lecture Hours:	Monday 12:30pm – 2:20pm (MW103) Thursday 12:30pm – 1:20pm (MW103)
Tutorial Hours:	Friday 1:30pm – 2:20pm (RR103)
Course Website:	http://moodle.hku.hk/ & http://www.statsoft.org/teaching/stat3622

Course Objectives

This course (as part of data science) will focus on explorative data analysis with statistical graphics and interactive data visualization. Students will learn how to display, communicate and analyze data, using a set of tools such as R and Python.

Programming: R (primary), Python, D3.js

Prerequisites:

STAT2602 Probability & Statistics II or STAT3902 Statistical Models

Contents and Topics

Data science, Data manipulation, Explorative data analysis, Statistical graphics, Interactive data visualization, other selected topics.

You are expected to learn:

- Choose the best chart that fits the data
- Communicate effectively using statistical graphics
- Create compelling visualization using computer software

You will: work hard and have fun!

Assessment Method

- **Exams:** None
- **Homeworks:** None

Assessment Method

- **Exams:** None
- **Homeworks:** None
- **Continuous:** Participation and in-class quizzes (40%)
- **Final Project:** Demo, oral presentaiton, written report (60%)

References

- Chang, W. (2013). *R Graphics Cookbook*. O'Reilly. <http://www.cookbook-r.com/>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd). Springer. <http://ggplot2.org/book/>; <http://hadley.nz/>
- Rossant, C. (2015). *Learning IPython for Interactive Computing and Data Visualization* (2nd). Packt. <http://ipython-books.github.io/minibook/>
- Murray, S. (2013). *Interactive Data Visualization for the Web: An Introduction to Designing with D3*. O'Reilly.
- Yau, N. (2011). *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Wiley. <http://book.flowingdata.com/>
- Ward, M. O., Grinstein, G. and Keim, D. (2015). *Interactive Data Visualization: Foundations, Techniques, and Applications* (2nd). CRC.

Online Materials

- D3 in Depth. <http://d3indepth.com/>
- Shiny Tutorial. <http://shiny.rstudio.com/tutorial/>
- RStudio Cheat Cheats. <https://www.rstudio.com/resources/cheatsheets/>

Data Visualization with ggplot2
Cheat Sheet

Studio

Basics

ggplot2 is a declarative system for creating graphics with the grammar of graphics. It is designed to be easy to learn and use, and to integrate well with the R ecosystem.

To create a plot, you need a data frame and a ggplot object. The data frame is a table of data, and the ggplot object is a collection of layers that define the plot's appearance.

Geoms

Geoms are the visual elements of a plot. They are created using the `geom_*` functions. The most common geoms are `geom_point()`, `geom_line()`, `geom_bar()`, and `geom_smooth()`.

Facets

Facets allow you to create multiple plots from the same data, based on one or more variables. They are created using the `facet_*()` functions.

Themes

Themes control the overall appearance of a plot, including the background, grid lines, and font. They are created using the `theme_*()` functions.

Annotations

Annotations are used to add text, lines, and shapes to a plot. They are created using the `annotate()` function.

Interactive Web Apps with shiny Cheat Sheet
Cheat Sheet

Studio

Basics

Shiny is a framework for building interactive web applications in R. It consists of a server and a client. The server runs R code that generates data and renders the user interface. The client is a web browser that displays the user interface and sends user input back to the server.

Building an App

Building an app involves several steps: creating a new app, writing the server and UI code, and running the app. The server code is written in R and the UI code is written in HTML and JavaScript.

Server Code

The server code is responsible for generating data and rendering the user interface. It is written in R and uses the `render_*` functions to output HTML, JavaScript, or other content.

UI Code

The UI code is responsible for creating the user interface. It is written in HTML and JavaScript and uses the `input_*` and `output_*` functions to create input and output elements.

Running an App

Running an app is done using the `runApp()` function. This function starts the Shiny server and opens a web browser to view the app.

Outline

- 1 Course Outline
- 2 Tentative Schedule**
- 3 RStudio and Markdown

Tentative Schedule

Subject to change, depending on time/class interests (frequent updates)

Part 1: Statistical Graphics		
Sep 1 (Th)	Introduction	Course Setup, RStudio
Sep 5 (M)	Statistics and Data Science	Start tutorials
Sep 8 (Th)	Explorative Data Analysis	
Sep 12 (M)	R Graphics I	Quiz 1
Sep 15 (Th)	R Graphics II	
Sep 19 (M)	ggplot2 I	
Sep 22 (Th)	ggplot2 II	
Sep 26 (M)	Time series visualization	
Sep 29 (Th)	Geospatial data visualization	
Part 2: Interactive Data Visualization		
Oct 3 (M)	Interactive Data Visualization	Quiz 2
Oct 6 (Th)	D3.js I	Call for Project Proposal
Oct 13 (Th)	D3.js II	
Oct 17-22	Reading Week	

Tentative Schedule

Part 2: Interactive Data Visualization (continued)		
Oct 24 (M)	rCharts, plotly	Quiz 3
Oct 27 (Th)	ECharts, Google Charts	
Oct 31 (M)	Shiny for App Building	
Part 3: Other Selected Topics		
Nov 3 (Th)	Text and Document	Quiz 4
Nov 7 (M)	Social Network	
Nov 10 (Th)	Model-based Cases	
Nov 14 (M)	Python, pandas, matplotlib	
Nov 17 (Th)	*High-dimensional data	
Nov 21 (M)	*Big data visualization	
Nov 24 (Th)	Oral Presentation I	Projects Portfolio
Nov 28 (M)	Oral Presentation II	

Outline

- 1 Course Outline
- 2 Tentative Schedule
- 3 RStudio and Markdown

RStudio IDE

- RStudio is a popular IDE¹/Interface for R.
- It is a powerful editor for R coding and debugging.
- It is a powerful generator for HTML, PDF, dynamic documents and slide shows.
- RStudio can be run on both Desktop and Server. In this course, we use RStudio Server on the cloud:
<http://stat3622.saas.hku.hk:8787>
You need not worry about software intallation.
- Check out more nice features of RStudio IDE at <https://www.rstudio.com/products/rstudio/features/>

¹IDE: Integrated Development Environment

R Markdown

Keywords: .Rmd Files, Dynamic documents, Reproducible research

[Click here to view a fantastic micro-video!](#)



[Click here for a gallery of creative works generated by R Markdown](#)