

# Week 1 Tutorial - STAT 485/685

**Trevor Thomson**

Department of Statistics & Actuarial Science  
Simon Fraser University, BC, Canada

September 14, 2020

SFU

# Today's Plan

## ① What is R?

- Advantages / Disadvantages
- How do I install R on my computer?

## ② What is RStudio?

- Advantages / Disadvantages
- How do I install RStudio on my computer?

## ③ Syntax and Basic Commands

## ④ Demonstration

## ⑤ Basic Concepts

- Example 1
- Example 2

SFU

# What is R?

- “*A language and environment for statistical computing and graphics*” - Kabacoff (2015).
  - Writes like a programming language:
    - if ( $x \leq 5$ ) {  $y = x + 2$  } else {  $y = x - 2$  }
    - while ( $i \in 1:n$ ) {  $y[i] = y[i-1] + rnorm(1)$  }
    - for ( $j \in 1:n$ ) {  $z[j] = z[j-1] + 2$  ;  $y[j] = z[j] / exp(pi)$  }
    - bellcurve = function( $x$ )  $1/sqrt(2*pi) * exp(-0.5 * x^2)$
  - Run computer programs by “sourcing” functions:
    - source(PenalizedLikFcn.R)
  - Create useful plots...
    - ...like the ones on the next slide!

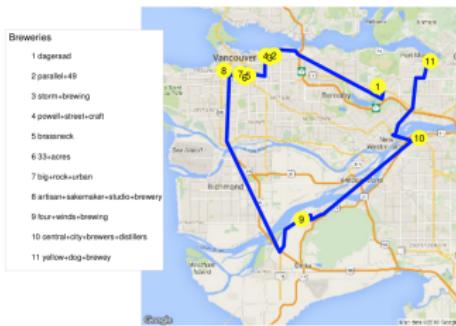
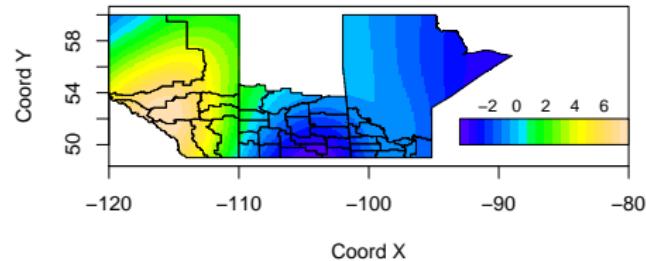


Figure 2: Route of our Brewery Trip.



# What is R?

## Advantages

### Why use R?:

1. R is 100% free!
2. R can virtually perform any statistical analysis
  - libraries are developed by the day!
3. R has state-of-the-art graphic capabilities.
4. R was developed for interactive data analysis and exploration.
  - Don't have to be an "expert" coder to use R!
5. R can easily import data from a wide variety of sources
  - includes: text, spreadsheets, and web data!
6. R runs on a wide array of platforms, including Windows, Unix, and Mac OS X.
7. If you don't want to learn a new language, a variety of graphic user interfaces are available.
8. R is open source!

SFU

# What is R?

## Disadvantages

### Why R causes a few headaches:

1. Quality of (some) packages are less than ideal.
  - Disadvantage of R being free!
2. Updates can result in functions / programs to break!
3. R was not designed to be *efficient!*
  - R *can* be slow
  - Disadvantage of R making programming easier.

# How do I install R?

**Step 1:** Google “R Cran” - go to <https://cran.r-project.org/>

The screenshot shows a Google search results page for the query "cran.r-project.org". The top navigation bar includes a search input with the query, a microphone icon, and a magnifying glass icon. Below the search bar are links for All, Images, News, Videos, Maps, More, Settings, and Tools. The search results indicate about 29,700,000 results found in 0.40 seconds. The first result is a link to the "The Comprehensive R Archive Network" at cran.r-project.org. The snippet for this result describes R as a freely available language and environment for statistical computing and graphics. Below the snippet are sections for Windows, Mirrors, and Available CRAN Packages for Mac OS X. To the right of the search results, there is a large thumbnail image of the R logo and a link to "More images". Further down the page, there is a summary of R as a programming language, its history, and its design by Ross Ihaka and Robert Gentleman. The page also mentions paradigms like Array programming, Functional programming, Procedural programming, and Reflection.

Google

cran.r-project.org

About 29,700,000 results (0.40 seconds)

[cran.r-project.org](https://cran.r-project.org/) ▾

**The Comprehensive R Archive Network**

What are R and CRAN? R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc.

**Windows**

R-4.0.2 for Windows (32/64 bit).  
Download R 4.0.2 for Windows ...

**(Mac) OS X**

R for Mac OS X. This directory contains binaries for a base ...

**Download R for Windows**

Binaries of contributed CRAN packages (for R >= 2.13.x ...)  
[More results from r-project.org »](#)

[www.r-project.org](http://www.r-project.org) ▾

**Mirrors**

CRAN Mirrors. The Comprehensive R Archive ...

**Available CRAN Packages for Mac OS X**

Available CRAN Packages By Name. A B C D E F G H I J K L ...

**Contributed Packages**

Contributed Packages. Available Packages. Currently, the CRAN ...

**R**

Programming language

R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. [Wikipedia](#)

**Designed by:** Ross Ihaka, Robert Gentleman

**Paradigms:** Array programming, Functional programming, Procedural programming, Reflection

# How do I install R?

## Step 2: Download R that is compatible with your operating system!



[CRAN](#)  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

[About R](#)  
[R Homepage](#)  
[The R Journal](#)

[Software](#)  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

[Documentation](#)  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-06-22, Taking Off Again) [R-4.0.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

### Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

# How do I install R?

**Step 3:** Click on “R-4.0.2.pkg” to start downloading R.



[CRAN  
Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

[About R](#)  
[R Homepage](#)  
[The R Journal](#)

[Software](#)  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

[Documentation](#)  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

## R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) [here](#). Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

Package binaries for R versions older than 3.2.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting (<https://cran-archive.r-project.org>) accordingly.

R 4.0.2 "Taking Off Again" released on 2020/06/22

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type

`md5 R-4.0.2.pkg`

in the Terminal application to print the MD5 checksum for the R-4.0.2.pkg image. On Mac OS X 10.7 and later you can also validate the signature using `pkgutil --check-signature R-4.0.2.pkg`

### Latest release:

[R-4.0.2.pkg](#) (notarized and signed)  
SHA1-hash: 7e4e1f0d407cd675e6eadd9f6a126ee9c83db3b  
(ca. 84MB)

R 4.0.2 binary for macOS 10.13 (High Sierra) and higher, signed and notarized package. Contains R 4.0.0 framework, R.app GUI 1.72 in 64-bit for Intel Macs, Tcl/Tk 8.6.6 X11 libraries and Texinfo 6.7. The latter two components are optional and can be omitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) to be installed since it is no longer part of OS X. Always re-install XQuartz when upgrading your macOS to a new major version.

**Important:** this release uses Xcode 10.1 and GNU Fortran 8.2. If you wish to compile R packages from sources, you will need to download and GNU Fortran 8.2 - see the [tools](#) directory.

[NEWS](#) (for Mac GUI)

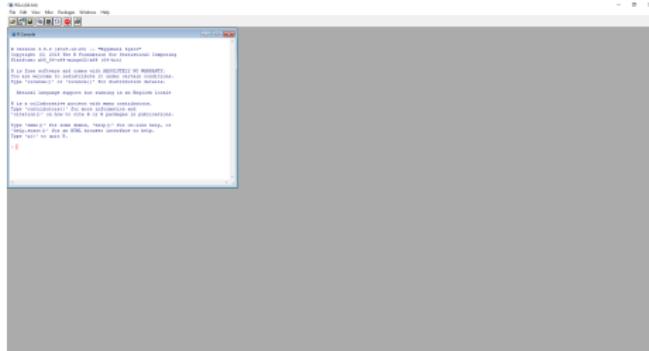
News features and changes in the R.app Mac GUI

[Mac-GUI 1.72 tar.gz](#)

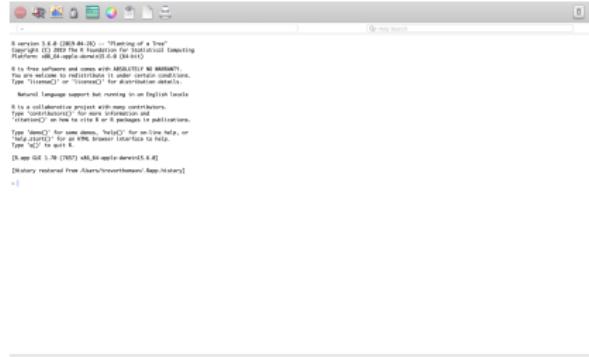
Sources for the R.app GUI 1.72 for Mac OS X. This file is only needed if you want to join the

# We Downloaded R!

Once R is downloaded, open it!



Windows



Mac OS X



# What is RStudio?

**Confession:** I don't know anyone that (directly) uses R!

# What is RStudio?

**Confession:** I don't know anyone that (directly) uses R!

*"Inspired by the innovations of R users in science, education, and industry, RStudio develops free and open tools for R and enterprise-ready professional products for teams to scale and share work. Our goal is to empower users to be productive with R."* - RStudio (2018)

# What is RStudio?

## Advantages

### Why use RStudio?:

1. RStudio is an add-on to R - still 100% free!
2. More “user-friendly”.
3. Better layout.
4. Easier to “debug”, view functions, examine data structures, and view plots.
5. Remembers your progress.
  - Useful to recover your code if your program crashes!

# What is RStudio?

## Disadvantages

### Why RStudio causes a few headaches:

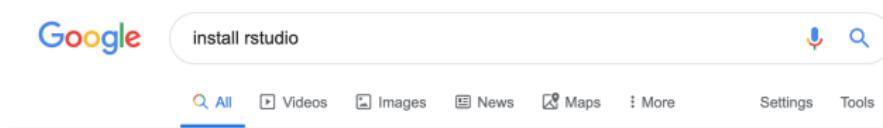
1. RStudio is an add-on to R - has the same issues as R!

SFU

# How do I install RStudio?

## Step 1: Google “Download Rstudio” - go to

<https://www.rstudio.com/products/rstudio/download/>



About 2,560,000 results (0.49 seconds)

### Download RStudio - RStudio

<https://www.rstudio.com/products/rstudio/download/> ▾

Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security policy. RStudio 1.2 requires ...

[RStudio - RStudio · RStudio Server · Import RStudio's public code ... · Release Notes](#)

### RStudio - RStudio

<https://www.rstudio.com/products/rstudio/> ▾

May 15, 2019 - Take control of your R code. RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. ... RStudio is available in open ...

### Download R and RStudio | UT.7.01x | edX

<https://courses.edx.org/courses/UTAustinX/.../56c5437b88fa43cf828bff5371c6a924/> ▾

To Install RStudio. Go to [www.rstudio.com](https://www.rstudio.com) and click on the "Download RStudio" button. Click on "Download RStudio Desktop." Click on the version recommended for your system, or the latest Windows version, and save the executable file. Run the .exe file and follow the installation instructions.

### Videos

# How do I install RStudio?

## Step 2: Scroll down to “Installers for Supported Platforms”

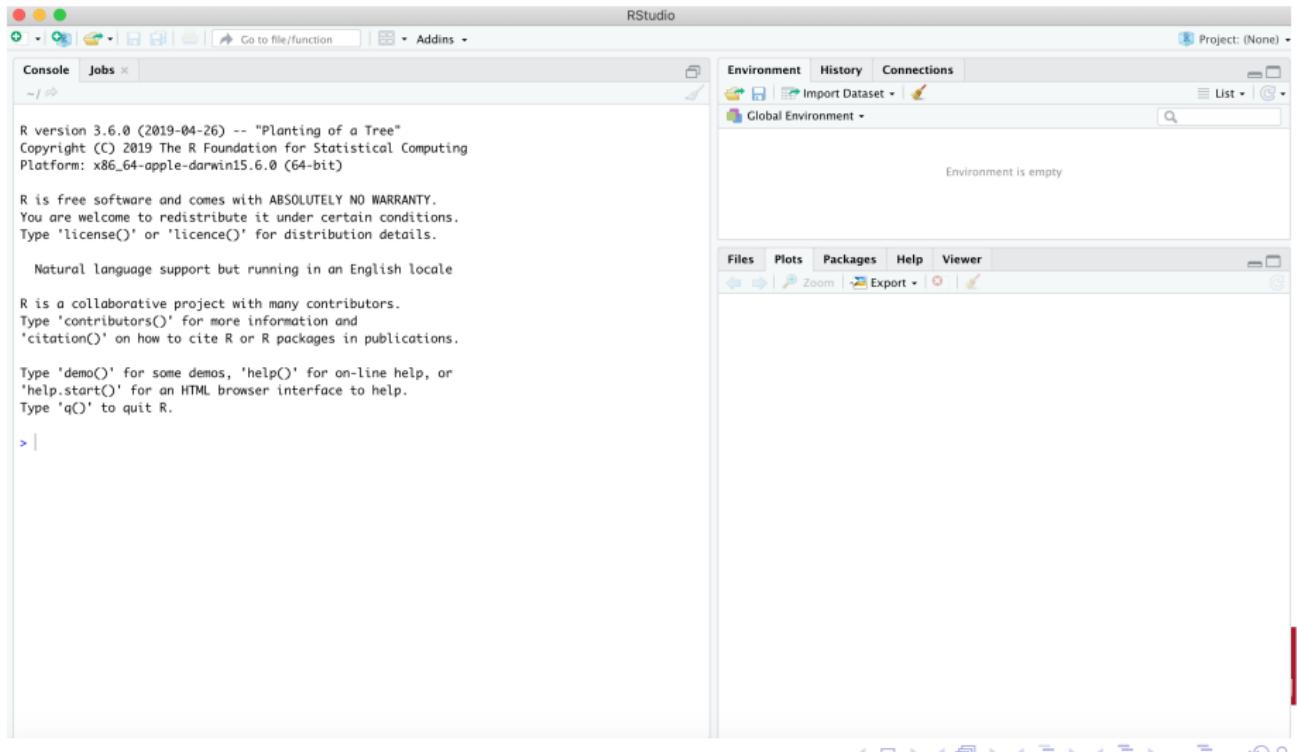
OS	Download	Size	SHA-256
Windows 10/8/7	<a href="#"> RStudio-1.3.1073.exe</a>	171.62 MB	2fea472a
macOS 10.13+	<a href="#"> RStudio-1.3.1073.dmg</a>	148.66 MB	0878b305
Ubuntu 16	<a href="#"> rstudio-1.3.1073-amd64.deb</a>	124.07 MB	6d71c5ff
Ubuntu 18/Debian 10	<a href="#"> rstudio-1.3.1073-amd64.deb</a>	126.78 MB	86be9352
Fedora 19/Red Hat 7	<a href="#"> rstudio-1.3.1073-x86_64.rpm</a>	146.95 MB	01abb3d8
Fedora 28/Red Hat 8	<a href="#"> rstudio-1.3.1073-x86_64.rpm</a>	151.04 MB	4b4e4878
Debian 9	<a href="#"> rstudio-1.3.1073-amd64.deb</a>	126.98 MB	0226bbc2
SLES/OpenSUSE 12	<a href="#"> rstudio-1.3.1073-x86_64.rpm</a>	119.43 MB	7c1a6f2c
OpenSUSE 15	<a href="#"> rstudio-1.3.1073-x86_64.rpm</a>	128.39 MB	29078f11

Zip/Tarballs



# We Downloaded RStudio!

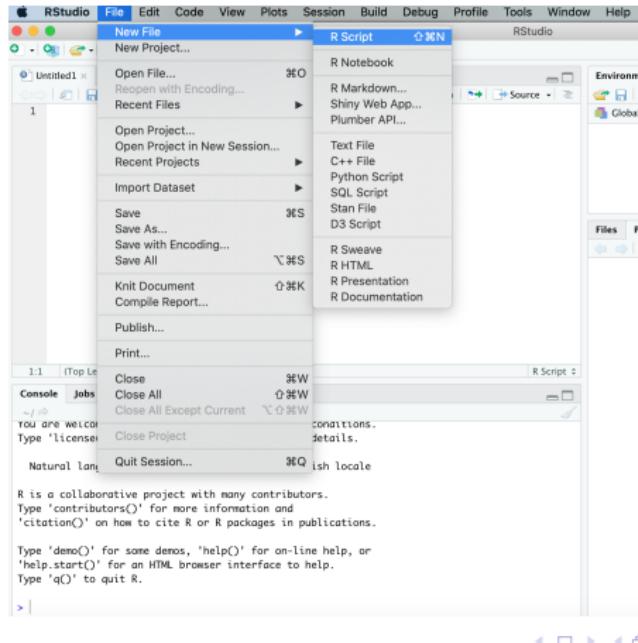
Once RStudio is downloaded, open it!



# We Downloaded RStudio!

Once RStudio is downloaded, open it!

Open up a new “R script” by pressing “Command + Shift + N” (for Mac users) or “Ctrl + Shift + N” (for Windows users)



# We Downloaded RStudio!

What are all these windows?

The screenshot shows the RStudio interface with several windows open:

- Script File**: A window titled "Untitled1" containing the text "This is where you write your code!".
- Environment / History**: A window showing the message "Environment is empty". It includes links to "See what objects are in your working space (Envirionment)" and "or view your command history (History)".
- Files / Plots / Packages / Help / Viewer**: A window showing the message "See file directories, view plots, see your packages, access "R Help", and view local web content".
- Console**: A window displaying R startup messages and a command prompt. It includes annotations:
  - "This is where your code from the Script file performs the computations"
  - "Also useful for simple calculations!"

# Syntax and Basic Commands

## Basic Flow of R code:

1. Define “inputs”
  - scalars, vectors, matrices, data frames, etc.
2. Define functions (or use already available functions in “packages”)
  - give them inputs, where they do “stuff” with them.
3. Do “stuff” with the output

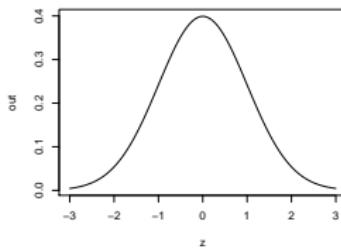
# Syntax and Basic Commands

## Basic Flow of R code:

1. Define “inputs”
  - scalars, vectors, matrices, data frames, etc.
2. Define functions (or use already available functions in “packages”)
  - give them inputs, where they do “stuff” with them.
3. Do “stuff” with the output

## Example (from earlier):

1. `z = seq(-3, 3, length.out=1001)`
2. `bellcurve = function(x) 1/sqrt(2*pi) * exp(-0.5 * x ^2)`  
`out = as.numeric() # initialize a vector to store results`  
`for (j in 1:length(z)){ out[j] = bellcurve(x=z[j]) }`
3. `plot(z, out, type="l")`



# Syntax and Basic Commands

## Common Operations in R:

For scalars  $x, y$

- Addition:  $x + y$
- Subtraction:  $x - y$
- Multiplication:  $x * y$
- Division:  $x / y$

For vectors  $x, y$

- Addition (component-wise):  $x + y$
- Subtraction (component-wise):  $x - y$
- Multiplication (component-wise):  $x * y$
- Division (component-wise):  $x / y$

For matrices  $X, Y$  (provided they are conformable)

- Addition:  $X + Y$
- Subtraction:  $X - Y$
- Multiplication:  $X \%*\% Y$
- Matrix Inversion (ie  $X^{-1}$ ): `solve(X)`
- Matrix transpose (ie  $X^t$ ): `t(X)`



# Syntax and Basic Commands

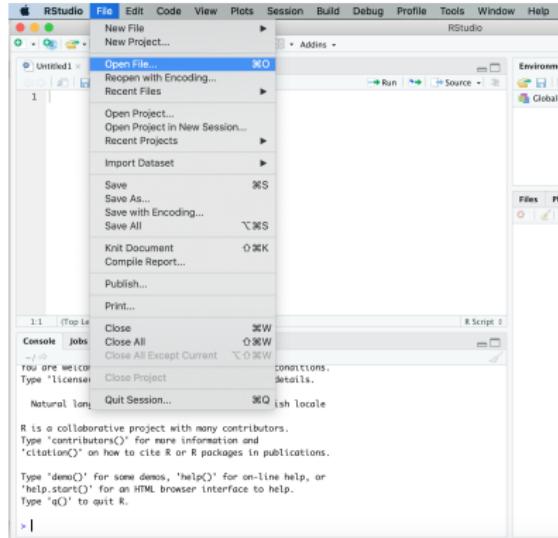
## Useful Functions in R:

For vectors  $x$ ,  $y$ , and matrix  $X$

Function	Description
<code>sum(x)</code>	Returns the sum of the elements of $x$
<code>prod(x)</code>	Returns the product of the elements of $x$
<code>min(x)</code>	Returns the minimum element of $x$
<code>max(x)</code>	Returns the maximum element of $x$
<code>mean(x)</code>	Returns the mean of the elements of $x$
<code>median(x)</code>	Returns the median of the elements of $x$
<code>sd(x)</code>	Returns the standard deviation of the elements of $x$
<code>var(x)</code>	Returns the variance of the elements of $x$
<code>cov(X)</code>	Returns the covariance between the columns of $X$
<code>cor(X)</code>	Returns the correlation between the columns of $X$
<code>sort(x, decreasing = FALSE)</code>	Returns the elements of $x$ sorted from smallest to largest
<code>unique(x)</code>	Returns the unique elements of $x$
<code>length(x)</code>	Returns the length of the vector $x$
<code>dim(X)</code>	Returns the dimension of the matrix $X$
<code>lm(y ~ x)</code>	Returns the line of best fit between $x$ and $y$
<code>plot(x, y)</code>	Returns a plot with $x$ and $y$ on the $x$ -axis and $y$ -axis, respectively
<code>c(x, y)</code>	Returns a vector with $x$ and $y$ as elements

# Demonstration

Open up an “R script” by pressing “Command + O” (for Mac users) or “Ctrl + O” (for Windows users)



We will open up the R Script Week1.R (available on Canvas!)

# Basic Concepts

- **Stochastic Process:** A collection of random variables indexed by some set (in our case, this set is  $\mathbb{Z}$  - the set of integers)
  - $\{Y_t : t \in \mathbb{Z}\}$
  - $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- **Mean function:**  $\mu_t = E(Y_t)$ , for  $t \in \mathbb{Z}$
- **Autocovariance function:**  $\gamma_{t,s} = Cov(Y_t, Y_s)$ , for  $t, s \in \mathbb{Z}$   
Note that  $\gamma_{t,t} = Var(Y_t)$ .
- **Autocorrelation function:**  $\rho_{t,s} = Corr(Y_t, Y_s) = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}} \times \sqrt{\gamma_{s,s}}}$ , for  $t, s \in \mathbb{Z}$

# Basic Concepts

## Example 1

Let  $e_1, e_2, \dots$  be a sequence of independent, identically distributed (iid) random variables, where  $E(e_t) = 0$ ,  $\text{Var}(e_t) = \sigma_e^2$ .

Let  $Y_t = Y_{t-1} + e_t$ , where  $Y_0 = 0$ . Note that  $\{Y_t : t = 1, 2, \dots\}$  is a stochastic process.

$$Y_1 = e_1$$

$$Y_2 = Y_1 + e_2 = e_1 + e_2$$

$$Y_3 = Y_2 + e_3 = e_1 + e_2 + e_3$$

⋮

$$Y_t = \sum_{u=1}^t e_u$$

SFU

**Question:** What is  $\mu_t$ ? What is  $\gamma_{t,s}$ ? What is  $\rho_{t,s}$ ?



# Basic Concepts

## Example 1

$\mu_t$ :

$$\mu_t = E(Y_t) = E\left(\sum_{u=1}^t e_u\right) = \sum_{u=1}^t E(e_u) = \sum_{u=1}^t 0 = 0$$

$\gamma_{t,s}$ :

$$\gamma_{t,t} = Var(Y_t) = Var\left(\sum_{u=1}^t e_u\right) = \underbrace{\sum_{u=1}^t Var(e_u)}_{\text{why?}} = \sum_{u=1}^t \sigma_e^2 = t\sigma_e^2$$

$$\begin{aligned}\gamma_{t,s} &= Cov(Y_t, Y_s) = Cov\left(\sum_{u=1}^t e_u, \sum_{v=1}^s e_v\right) \\ &= \sum_{u=1}^t \sum_{v=1}^s Cov(e_u, e_v) = \min\{t, s\}\sigma_e^2 + 0 = \min\{t, s\}\sigma_e^2\end{aligned}$$

$\rho_{t,s}$ :

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}} \times \sqrt{\gamma_{s,s}}} = \frac{\min\{t, s\}\sigma_e^2}{\sqrt{t\sigma_e^2} \times \sqrt{s\sigma_e^2}} = \begin{cases} \sqrt{\frac{t}{s}} & \text{if } 1 \leq t \leq s \\ \sqrt{\frac{s}{t}} & \text{if } 1 \leq s \leq t \end{cases}$$

What can we say about  $\rho_{t,s}$ ?

# Basic Concepts

## Example 2

Let  $\dots, e_{-2}, e_{-1}, e_0, e_1, e_2, \dots$  be a sequence of independent, identically distributed (iid) random variables, where  $E(e_t) = 0$ ,  $Var(e_t) = \sigma_e^2$ .

Let  $Y_t = \frac{e_t + e_{t-1}}{2}$ . Note that  $\{Y_t : t \in \mathbb{Z}\}$  is a stochastic process.

**Question:** What is  $\mu_t$ ? What is  $\gamma_{t,s}$ ? What is  $\rho_{t,s}$ ?

# Basic Concepts

## Example 2

$\mu_t$ :

$$\mu_t = E(Y_t) = E\left(\frac{e_t + e_{t-1}}{2}\right) = \frac{E(e_t) + E(e_{t-1})}{2} = \frac{0 + 0}{2} = 0$$

$\gamma_{t,s}$ :

$$\gamma_{t,t} = \text{Var}(Y_t) = \text{Var}\left(\frac{e_t + e_{t-1}}{2}\right) \underset{\text{why?}}{=} \frac{\text{Var}(e_t) + \text{Var}(e_{t-1})}{4} = \frac{\sigma_e^2 + \sigma_e^2}{4} = \frac{\sigma_e^2}{2}$$

$$\gamma_{t,s} = \text{Cov}(Y_t, Y_s) = \text{Cov}\left(\frac{e_t + e_{t-1}}{2}, \frac{e_s + e_{s-1}}{2}\right) = \frac{1}{4} \text{Cov}(e_t + e_{t-1}, e_s + e_{s-1})$$

$$= \frac{1}{4} [\text{Cov}(e_t, e_s) + \text{Cov}(e_t, e_{s-1}) + \text{Cov}(e_{t-1}, e_s) + \text{Cov}(e_{t-1}, e_{s-1})]$$

$$= \begin{cases} 1/2 \times \sigma_e^2 & \text{if } |t - s| = 0 \\ 1/4 \times \sigma_e^2 & \text{if } |t - s| = 1 \\ 0 & \text{otherwise} \end{cases}$$

(Continued on next slide)

SFU

# Basic Concepts

## Example 2

$$\gamma_{t,s} = \frac{1}{4} [Cov(e_t, e_s) + Cov(e_t, e_{s-1}) + Cov(e_{t-1}, e_s) + Cov(e_{t-1}, e_{s-1})]$$

$$\gamma_{t,t-1} = \frac{1}{4} [Cov(e_t, e_{t-1}) + Cov(e_t, e_{t-2}) + Cov(e_{t-1}, e_{t-1}) + Cov(e_{t-1}, e_{t-2})]$$

$$= \frac{1}{4} [0 + 0 + \sigma_e^2 + 0] = \frac{1}{4} \sigma_e^2$$

$$\gamma_{t-1,t} = \underbrace{\frac{1}{4} \sigma_e^2}_{(*)}$$

$$\gamma_{t,t-j} = \frac{1}{4} [Cov(e_t, e_{t-j}) + Cov(e_t, e_{t-j-1}) + Cov(e_{t-1}, e_{t-j}) + Cov(e_{t-1}, e_{t-j-1})]$$

$$= \frac{1}{4} [0 + 0 + 0 + 0] = 0, \text{ for } j \geq 2$$

$$\gamma_{t-j,t} = 0, \text{ for } j \geq 2$$

(\*) **Exercise:** Can you show this without using properties of covariance?

$\rho_{t,s}$ :

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}} \times \sqrt{\gamma_{s,s}}} = \begin{cases} 1 & \text{if } |t-s| = 0 \\ 1/2 & \text{if } |t-s| = 1 \\ 0 & \text{otherwise} \end{cases}$$

