

DSCI 351-351M-451 Syllabus

Exploratory Data Science

Fall 2022 Tuesday, Thursday 11:30 am to 12:45 pm NORD 356 &
Remote/Synchronous

Prof. Roger H. French (CWRU),
Prof. Laura S. Bruckman (CWRU),
Prof. Paul Leu (Pitt),
Prof. Kris Davis (UCF),
Prof. Sonya Cirlos (UTRGV)

November 29, 2022

Contents

1	Joint Undergraduate and Graduate Course	2
1.1	Case Western Reserve University (CWRU): DSCI351 and DSCI351M	2
1.2	Case Western Reserve University (CWRU): DSCI451	2
1.3	Univ. of Pittsburgh (Pitt): ENGR 1451 and ENGR 2451	3
1.4	Univ. of Central Florida (UCF): EMA 6908	3
1.5	Univ. of Texas, Rio Grande Valley (UTRGV): CSCI 4341-04	4
2	Course Description	4
2.1	Class Repository Folder Structure	5
2.2	License applied to course materials and some datasets used for data analysis	5
2.3	Outcomes	6
2.4	Prerequisites	6
3	Textbooks and Readings	6
4	Grading: Lab Exer., SemProj Rep. & Pres., MidTerm & Final Exam	8
5	DSCI351-351M-451 Syllabus: Weekly Topics	8
6	Contact Information	10
7	Course Mechanics	11
7.1	Lectures	11
7.2	Consultations	11
7.3	Lab Exercise Assignments	11
7.4	451 Semester Project Template	11
7.5	CWRU Data Science Slack	12
8	Coding and Data Science Tools and Resources	12

9 Policies	14
9.1 Attendance	14
9.2 Readings	14
9.3 Lab Exercise Assignments	14
9.4 Collaboration and Citation	15
9.5 Academic Integrity Policy	15
9.6 Disability Resources	15
10 Copyleft, References, Citations & Rubrics	15
10.1 CopyLeft	15
11 Setting up your R data science computer	16
11.1 R Standard Packages for Win,Lin,Mac	16
11.2 For Windows	17
11.2.1 LaTeX	17
11.2.2 Git	21
11.2.3 R	27
11.2.4 GVim	30
11.3 For Linux	30
11.3.1 LaTeX	30
11.3.2 Git	30
11.3.3 R	31
11.4 For Mac	33
11.4.1 Homebrew	33
11.4.2 XQuartz	33
11.4.3 LaTeX	34
11.4.4 Git	34
11.4.5 R	34
References	34

1 Joint Undergraduate and Graduate Course

1.1 Case Western Reserve University (CWRU): DSCI351 and DSCI351M

DSCI351 is the 3rd level class in the Applied Data Science Undergraduate Minor. The ADS Minor is available to CWRU students across all the schools in the University. For more information see Applied Data Science Undergraduate Minor in the CWRU Bulletin [Undergraduate minor in Applied Data Science](#).

DSCI351 will introduce students to the basic elements of a data analysis, including R coding, Rstudio IDE and Git version control, statistical concepts, the stages of a data analysis and reproducible research.

DSCI351M section focuses specifically on Exploratory Data Science for Materials and Materials' Systems.

1.2 Case Western Reserve University (CWRU): DSCI451

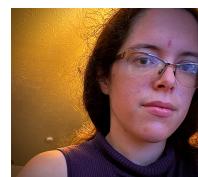
DSCI 451 is a graduate level introduction to data science and analytics. DSCI 451 is one of the classes that satisfy the University Graduate Certificate in Applied Data Science. For more information on the University Graduate Certificate in Applied Data Science, see [Grad. Cert. in Applied Data Science](#) that is now available to students in all CWRU Schools and Colleges. Graduate students will develop a semester long data science project focused on time-series, spectral, image, statistical process control or machine learning data science problems that are relevant to their research interests. These projects will include preparing datasets, code scripts and functions, a git repository for other students to use these codes as open source resources, and the preparation of reproducible data science analyses for these problems.



(a) CWRU Prof. Roger French.



(b) CWRU TA: Will Oltjen



(c) CWRU TA: Kristen Hernandez

Figure 1: The CWRU Team.

1.3 Univ. of Pittsburgh (Pitt): ENGR 1451 and ENGR 2451

As part of a research collaboration between CWRU and Pitt, where we have established the [NSF IUCRC Center for Materials Data Science for Reliability and Degradation](#), this year we are co-teaching the UG and GS DSCI 351/451 and DSCI 353/453 classes so that students at Pitt will learn data science with us this semester. The Pitt students will submit their Lab Exercises and Exams to the Pitt Canvas site for ENGR 1451/2451, and these will be graded independently of the CWRU students.



(a) UCF Prof. Paul Leu.



(b) UCF TA: Mingxuan Li

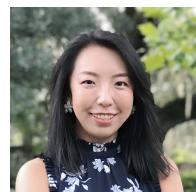
Figure 2: The Pitt Team.

1.4 Univ. of Central Florida (UCF): EMA 6908

As part of a research collaborations between CWRU and UCF, we have studied photovoltaics and reliability of technologies under real-world exposure conditions. We are co-teaching the UG and GS DSCI 351/451 and DSCI 353/453 classes so that students at UCF will also be able to learn data science with us this semester. The UCF students will submit their Lab Exercises and Exams to the UCF Canvas site for EMA 6908, and these will be graded independently of the CWRU students.



(a) UCF Prof. Kris Davis.



(b) UCF TA: Mengjie Li



(c) UCF TA: Dylan Colvin

Figure 3: The UCF Team.

1.5 Univ. of Texas, Rio Grande Valley (UTRGV): CSCI 4341-04

As part of ongoing collaborations between CWRU and UCF, we are co-teaching the UG and GS DSCI 351/451 and DSCI 353/453 classes so that students at UTRGV will also be able to learn data science with us this semester. The UTRGV students will submit their Lab Exercises and Exams to the UTRGV Canvas site for CSCI 4341-04, and these will be graded independently of the CWRU students.



(a) UTRGV Prof. Sonya Cirlos



(b) UTRGV Prof. Emmett Tomai

Figure 4: The UTRGV Team.

So lets welcome our Pitt, UCF and UTRGV partners!

2 Course Description

In this course, we will learn data science and analysis approaches to identify statistically significant relationships and better model and predict the behavior of these systems. We will assembly and explore real-world datasets, perform clustering and pair-wise correlation plot analyses to investigate predictor/variable correlations, and linear and logistic regression will be employed to develop associated predictive models. Results will be interpreted, visualized and discussed.

We will introduce basic elements of statistical analysis using the R coding language ([R Project open source software](#)) for exploratory data analysis and model development. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and munging functions, and a rich selection of statistical packages, used for data analytics, model development and prediction. This will include an introduction to R data types, reading and writing data, looping, plotting and regular expressions, so that one can start performing variable transformations for linear fitting and developing structural equation models, while exploring for statistically significant relationships. We will also learn tidy principles for data analysis, pipes and ggplot vs base graphics approaches to data visualization. Python is also commonly used for data science and data analyses, while at the same time Python is a general purpose programming language. Both R and Python are interpreted languages that do not require compiling the code prior to execution. Due to Python's broader spread of use cases (from data analysis to full applications to software engineering) it can be easier for a developing data scientist to find useful answers to questions, by learning R first. Python can then be learned as a second data analysis language. Students are welcome to use Python in the DSCI classes.

R Analytics will be applied to the case of time-series, spectral, image problems using continuous and categorical variables and datasets to analyze system responses, combined with results of experiments to identify fundamental principles that are statistically significant in the observed system performance. We will learn about longitudinal and cross-sectional data science studies, along with retrospective and prospective studies.

The class is taught using a “practicum” approach and will be structured to have a balance of theory and practice. We'll split class into Foundation and Practicum a) Foundation: lectures, presentations, discussion b) Practicum: coding, demonstrations and hands-on data science work.

Every student will have access to their own pre-configured Open Data Science Desktop computer, already configured for fast and easy adoption of good data science practices and tools. And the ODS VDIs are updated each month so as to have the latest versions of R packages and software for data science. And also introduces students to doing applied data science and data analysis on both Windows or Macintosh notebook computers, and also in Linux based High Performance Computing environment, using Markov, CWRU's data science cluster.

2.1 Class Repository Folder Structure

Please browse within each folder to learn more about the intended purpose of each folder in the standard structure.

This folder structure has been designed to accommodate each type of file you may need to create and modify - please do not create additional folders in the structure, and please pay attention to naming conventions when creating new files.

Course Material Folders in this Course Repository

- 1-Assessments is where you will find the Lab Exercises and Exams
- 2-Readings folder contains textbooks and readings for the course
- 3-Class contains daily class notes as *.Rmd and *.pdf files
- These are split into a Foundations "f" and a Practicum "p" class notes
- 4-Syllabus contains the updated Course Syllabus
- 5-SemProj contains the materials needed for DSCI451

Your Working Data Analysis Folders

- Scripts is where to write your scripts for data analysis
- Data contains course datasets and your datasets
- Figs is the figures folder, accessible for both Scripts, Topics and Docs
- Topics is where to write reports and presentation for your data analysis
- Docs is where to write formal documentation as *.Rmd, *.tex files
- Packages is where to build R packages, if your project involves this

2.2 License applied to course materials and some datasets used for data analysis

Class materials

- License: This work is legally bound by the following software license: [CC-A-NS-SA-4.0][1]
- Please see the LICENSE.txt file, in the root of this repository, for further details.

Assessment materials

- Lab Exercise Assignments, Project Assignments and Exams are all rights reserved.
- They are NOT creative commons licensed, and can not be distributed.

Datasets derived from funded research projects

- During this class you may be working on a project that is part of a funded research award at Case Western Reserve University.
- Information or material made available to you in connection with this funded research project, and coursework, data, results or other intellectual property you may develop in conjunction with this project, will be subject to Case Western Reserve's Intellectual Property Policy as well as terms of the sponsored research agreement.
- You acknowledge that you understand that you will not have ownership of intellectual property created in conjunction with the project.
- Please sign the "2208-DSCI-Acknowledgement-of-IP.pdf" form. And email it to Jonathan Steirer at jws227@case.edu

2.3 Outcomes

Capabilities

- Familiarity with R Statistics, scripting, functions, packages, automated data analysis.
- Familiarity with exploratory data analysis, statistical model building.
- Familiarity with Principles of Tidy Data Science from the tidyverse package and use of pipes (`%>%`) from the magrittr package.
- Applications of domain knowledge and statistical analytics to identify important predictors and develop initial predictive models.
- Applications of domain knowledge and analytics to identify important predictors and develop initial predictive models.
- Introduction to methods of reproducible research, including markdown, LaTeX and Git.

Data types include:

- Time-series, spectral, image and higher order datatypes,
And their assembly to produce augmented and derivative datasets.

Data set characteristics will include:

- Variety: Types of data and information, including both structured and unstructured data.
- Volume: Data from human sources (vendors, suppliers, distributors, customers, etc.) and sensor networks, both small and large data volumes.
- Velocity: Short time interval datasets.

2.4 Prerequisites

1. ENGR131 Elementary Computer Programming, EECS132, or equivalent
2. STAT312R Basic Statistics for Engineering and Science or equivalent

3 Textbooks and Readings

Required Texts and their Abbreviation, which is used in the syllabus:

Peng R Programming (PRP) and Peng Exploratory Data Analysis (EDA) are introductory books to R and Data Science and Analysis. These are LeanPub books, available from LeanPub for a "pay what you want" price.

R for Data Science (R4DS) is a new book teaching R for Tidy Data Science, and is available as a bookdown book on the web [R for Data Science](#), you can buy an ebook for \$18 from [Google Play Store page for this book](#).

Open Intro Statistics version 4 (OIS) is an open source text book on Inferential Statistics, published under a Creative Commons license, [for free distribution as a pdf](#). In addition a copy can be purchased from Amazon for \$20.

Introduction to Statistical Learning with R, 2nd Edition (ISLR) was just released this year and is a Springer book which is also available for [free as a pdf](#). ISLR is the text book used extensively in DSCI353/453 on Statistical and Machine Learning.

Additional reading assignments will be distributed via the course git repository in the readings subdirectory.

R Programming for Data Science



Roger D. Peng

Figure 5: **PRP:**
Roger Peng, **R**
Programming for
Data Science. 2014
[1]

Exploratory Data Analysis with R



Roger D. Peng

Figure 6: **EDA:**
Roger Peng, **Ex-**
ploratory Data
Analysis With R.
2015 [2]

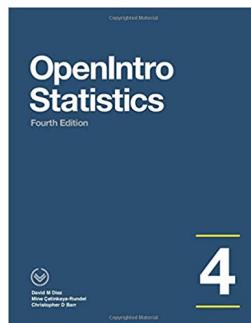


Figure 7: **OIS:** David
M. Diez, Christopher
D. Barr, and Mine
Cetinkaya-Rundel,
OpenIntro Statistics 4th Ed. 2015
[3]

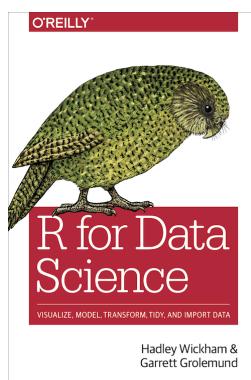


Figure 8: **R4DS:**
Garrett Grolemund,
Hadley Wickham **R**
for Data Science.
2017 [4]

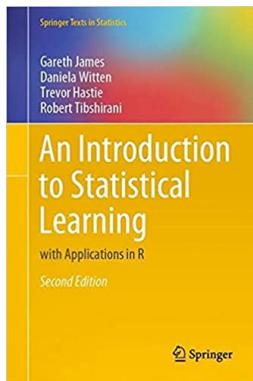


Figure 9: **ISLR:**
Gareth James,
Daniela Witten,
Trevor Hastie,
Robert Tibshirani
An Introduction to Statistical Learning: with Applications in R,
2nd Edition, 2021
[5]

4 Grading: Lab Exer., SemProj Rep. & Pres., MidTerm & Final Exam

DSCI351 and DSCI351M is graded on 100 points basis

Seven Lab Exercise (LE) Data Analyses,

LE1, LE2 = 7 points each, = 14 pts.

LE3 – LE7 worth 10 points each = 50 pts.

Three SemProject Report Presentation Evaluations (2pt/each) = 6 pts.

Midterm Exam = 10 pts.

Final Exam = 20 pts.

Total = 100 pts

DSCI451 is graded on a 140 point basis

Seven Lab Exercise (LE) Data Analyses,

LE1, LE2 = 7 points each, = 14 pts.

LE3 – LE7 worth 10 points each = 50 pts.

Midterm Exam = 10 pts.

Final Exam = 20 pts.

Three Report Presentation Evaluations (2pt/each) = 6 pts.

Six SemProj Updates worth (2point each) = 12 pts.

One SemProj with 3 presentations(5pts/each) & 1 Final Report(13pts) = 28 pts.

Total = 140pts

Grading is done on a curve.

All Lab Exercise assignments are submitted electronically through Canvas, uploading to the LE assignment page. by Midnight on the day they are due.

Late submission of Lab Exercises will result in loss of 1 point for each day they are late.

More details about grading are in the Course Mechanics section of the syllabus.

5 DSCI351-351M-451 Syllabus: Weekly Topics

Day:Date	Foundation	Practicum	Reading	Due
w01a:Tu:8/30/22	ODS Tool Chain	R, Rstudio, Git		
w01b:Th:9/1/22	Setup ODS Tool Chain	Bash, Git, Slack, Agile	PRP4-33	LE1
w02a:Tu:9/6/22	Bash-Git-Knuth-Lit.Prog.	RIntroR	PRP35-64	
w02b:Th:9/8/22	What is Data Science	OIS:Intro2R	OIS1,2	
w02Pr:Fr:9/9/22			PRP65-93	451 Update1
w03a:Tu:9/13/22	Data Intro	Data Analytic Style	PRP94-116	LE2 LE1 Due
w03b:Th:9/15/22	Rand. Var. Normal Dist.	Git, Rmds, Loops	OIS4	
w04a:Tu:9/20/22	Tidy Check Explore	Tidy GapMinder	EDA1-31	
w04b:Th:9/22/22	Inference, DSCI Process	Other Distrib. 7 ways	R4DS1-3	LE3 LE2 Due
w04Pr:Fr:9/23/22			EDA32-58	451 Update2
w05a:Tu:9/27/22	OIS4 Rand. Var.	EDA of PET Degr.	OIS5	
w05b:Th:9/29/22	OIS5 Found. of Infer.	Multivar Corr. Plot	R4DS4-6	
w05Pr:Fr:9/30/22				451 RepOut1
w06a:Tu:10/4/22	Pred., Algorithm, Model		R4DS7-8	
w06b:Th:10/6/22	Summ. Stats & Vis.	Anscombe's Quartets	R4DS9-16	LE4 LE3 Due
w06Pr:Fr:10/7/22				451 Update3
w07a:Tu:10/11/22	Midterm Rev. Tidy Data	Correl Plots Summ Stats	OIS6.1-2	PeerRv1 Due
w07b:Th:10/13/22	HypoTest, Infer. Recap	Penguin EDA, Sampling		
w08a:Tu:10/18/22	MIDTERM	EXAM		
w08b:Th:10/20/22	Programming & Coding	Code Packaging		LE4 Due
w08Pr:Fr:10/21/22				451 Update4
Tu:10/24,25	CWRU	FALL BREAK	R4DS17-21	
w09b:Th:10/27/22	Cat. Inf. 1 & 2 propor.	Indep. Test,2-way tables	OIS6.3-4	LE5
w09Pr:Fr:10/28/22				451 RepOut2
w10a:Tu:11/1/22	Goodness of Fit, χ^2 test	t-tests 1&2 means	OIS7.1-4	
w10b:Th:11/3/22	Num. Infer, Cont. Tables	Stat. Power		
w10Pr:Fr:11/4/22				451 Update5
w11a:Tu:11/8/22	Sample & Effect Size	Stat. Power GGmap	OIS8	PeerRv2 Due
w11b:Th:11/10/22	Regr Part 1, Test & Train	Curse of Dimen.	ISLR1,2.1,2	LE6 LE5 Due
w12a:Tu:11/15/22	Regr. Outliers	Regr Part 2, GIS	OIS9	
w12b:Th:11/17/22	Mult.Regr., Var. Select	Regr. Diagnostics		
w12Pr:Fr:11/18/22				451 Update6
w13a:Tu:11/22/22	Log. Regr.	Mult. Regression	ISLR3.1	LE7 LE6 due
w13b:Th:11/24/22	Statistical learning	Logistic Regr.	ISLR3.2	
w13Pr:Fr:11/25/22				451 RepOut3
w14a:Tu:11/23/22		GIS Trends	ISLR4.1-3	
Th,Fr:11/24,25	THANKSGIVIING	Vacation		
w15a:Tu:11/29/22	Classificat., Sup. Lrning	Log. Regr. & ML		PeerRv3 Due
w15b:Th:12/1/22	Clustering, Unsup. Lrn-ing	Caret, Broom 4 modeling	Fr.Br.2020	
w15SPr:Fr:12/2/22				
w16a:Tu:12/6/22	Big Data Analytics	Dist. Comp., Hadoop	Khalil.2020	
w16b:Th:12/8/22	Final Exam Review		Mirletz,2015	LE7 due
Friday 12/12	SemProj	Final Report		SemProj4 due
Monday 12/19	FINAL EXAM	12:00-3:00pm	Nord 356	or remote

Table 1: DSCI351-451 Weekly Syllabus. w01a is week 1, class a. w01b is week 1 class b. w02Pr is DSCI451 SemProj. Readings are defined by book and chapters, sections in Peng R Prog. (PRPx.y), Peng Exp. Data An. (EDAx.y), R for Data Sci. (R4DSx.y), Open Intro Stats (OISx.y) & Intro. to Stat. Learn. with R (ISLRx.y). 29, 2022

6 Contact Information

Prof. Roger H. French

- White 536, and electronically.
- Email is best: rxf131@case.edu, Use DSCI351 in the subject line
- @frenchrh on twitter
- Office Phone 216 368 3655, Cell Phone 302 468 6667

DSCI 351-351M TA: Will Oltjen

- White 614, and electronically.
- Email is best: wco3@case.edu, Use DSCI351-451 in the subject line
- @WillOltjen on twitter

DSCI 351-351M TA: Kristen Hernandez

- White 615, and electronically.
- Email is best: kjh125@case.edu, Use DSCI351-451 in the subject line

DSCI 451 Prof. Laura S. Bruckman

- White 502A, and electronically.
- Email is best: lsh41@case.edu, Use DSCI351 in the subject line
- @bruckman_laura on twitter

DSCI 451 TA: Nat Tomczak

- White 615, and electronically.
- Email is best: nkt8@case.edu, Use DSCI351-451 in the subject line

ENGR 1451-2451 Pitt Prof: Paul Leu

- electronically.
- Email is best: pleu@pitt.edu, Use ENGR 1451-2451 in the subject line
- @Paul_W_Leu on twitter

ENGR 1451-2451 Pitt TA: Mingxuan Li

- electronically.
- Email is best: MIL152@pitt.edu or mxl1261@case.edu, Use ENGR 1451-2451 in the subject line

EMA 6908 UCF Prof: Kris Davis

- electronically.
- Email is best: kristopher.davis@ucf.edu, Use EMA 6908 in the subject line
- @kris_o_davis on twitter

EMA 6908 UCF TA: Mengjie Li

- electronically.
- Email is best: menglie.li@ucf.edu, Use EMA 6908 in the subject line

EMA 6908 UCF Prof: Dylan Colvin

- electronically.
- Email is best: kristopher.davis@ucf.edu, Use EMA 6908 in the subject line
- @kris_o_davis on twitter

CSCI 4341-04 Prof: Sonya Cirlos

- electronically.
- Email is best: sonya.cirlos01@utrgv.edu, Use CSCI 4341-04 in the subject line
- @CirlosSonya on twitter

7 Course Mechanics

7.1 Lectures

Fall 2012 Tuesday, Thursday 11:30 am to 12:45 pm, Nord 356 and Remote/Synchronous on Zoom

7.2 Consultations

Office hours are Monday and Wednesday from 4pm to 5pm on Zoom, invites in the Canvas Course Calendar.
Lab Exercise questions and solutions can be discussed in Office Hours or in the class Slack Channel.
After class or as needed. Contact Prof. French or the TAs, by Slack, email.

7.3 Lab Exercise Assignments

All Lab Exercise assignments are submitted electronically through Canvas, uploading to the LE assignment page. by Midnight on the day they are due.

Late submission of Lab Exercises will result in loss of 1 point for each day they are late.

Filenames should contain DSCI351, -LE#, -YourLastName... e.g. DSCI351-LE2-French.R or DSCI351-LE2-French.R and DSCI351-LE2-French.pdf.

Lab Exercises need to be legible, organized and explain your thinking, process and results. Credit all resources you drew upon, including texts, papers, peers.

Lab Exercises are due prior to class on the day they are due.

Lab Exercises will be graded on canvas and reviewed in class.

Lab Exercise solutions will be reviewed in Office Hours

7.4 451 Semester Project Template

Students in DSCI 451 and ENGR 2451 have a 40 point semester-long data science project to do.

Prof. Laura Bruckman and TA Raymond Wieser organize and consult on the semester projects.

Please see the projectreport-template.Rmd for the template for 451 Updates and Report Outs Report Outs can be presented as the projectreport-template or as an .Rpres presentation

7.5 CWRU Data Science Slack

We have a CWRU Data Science Slack

You can join by signing up at <https://cwrudsci.slack.com>

Sign up using your case.edu email address

There is a specific class channel, along with general, applied data science and job opportunities channels.

The class Slack channel is a good place for discussing questions from class and/or assignments.

8 Coding and Data Science Tools and Resources

Markov Data Science Cluster

You will not need to install software on your personal computers.

We have a High Performance Computing Cluster dedicated to Data Science courses.

It is named Markov after [Andrey Markov](#).

To access Markov, sign in to the OnDemand website using your caseID, <https://ondemand.case.edu>

There you will see the an icon for the OnDemand Apps.

The “RStudio Server R/4.1.1, Shared by Roger French (rxf131)” is what we will use for class.

As a back up, you will also see “RStudio Server R/3.6.3, Shared by Roger French (rxf131)” which is using the prior version of the R programming language.

It can be useful to login using the Firefox or Vivaldi browwers, since OnDemand wants you to quit your browser at the end of a session.

Open Data Science (ODS) Win10 Desktops

You will not need to install software on your personal computers.

Instead you can login to the CWRU My Apps portal. [6]

The CWRU MyApps portal is located at <https://myapps.case.edu>

How to access your ODS Desktop from a computer

The <https://myapps.case.edu>, allows us to better scale back-end processing as necessary based on load.

All apps and student files have been migrated.

There will be more public announcements of MyApps in the coming weeks, but we wanted to make sure that you had the information now.

You can access your ODS Desktop, from any modern Browser (Chrome, Firefox).

Or you can install the Citrix Workspace Client program on you computer

-
- Documentation is included below.
 - [”MyApps Streamed Applications and Desktops - Usage Manual”](#)
 - Windows, Linux
 - [MyApps Citrix Workspace Setup \(Windows\)](#)
 - [Reset Citrix Workspace App on Windows](#)
 - macOS:
 - [MyApps Citrix Workspace Setup \(macOS\)](#)
 - [Reset Citrix Workspace App on macOS](#)

If there are any issues accessing MyApps, please call the HelpDesk (368-HELP, help@case.edu to submit a ticket, and CSE-IT team will assist.

Scripting, Coding and Writing

And more resources for open science coding and scripting, including tools for code editing, code version control and languages.

R Statistics

We will be using R in this class for Lab Exercises and SemProjs. Its generally useful language for statistical analysis and data science.

- [The R Project for Statistical Computing](#) [7] main website
- [R programming language](#) R is a free software programming language and software environment for statistical computing and graphics.[8]
- [RStudio](#) provides popular open source and enterprise-ready professional software for the R statistical computing environment. [9]
- [Google's R Style Guide](#)

Rmarkdown as a path to open access and reproducible science

- [R Markdown — Dynamic Documents for R](#). We will be doing all our work using Rmarkdown this semester. Class presentations, Lab Exercises, semester projects, all done in Rmd, as reproducible science projects, including data, code, and final output.
- [Introduction to R Markdown](#).
- [R Markdown Cheat Sheets](#).
- [An Rmarkdown Introduction](#) slideshow done from Rmarkdown and shared publicly on RPubs.

R Statistics, more resources

We will be using R in this class for lab exercises and projects. Its generally useful language for statistical analysis and data science.

- [The R Project for Statistical Computing](#) [7] main website
- [Roger Peng's Computing for Data Analysis introduction to R Statistics](#). These are from a Coursera course he does, with the same name. [10]
- [A \(very\) short introduction to R](#) [11]
- [Google's R Style Guide](#)
- [Hadley Wickham's R Style Guide](#)
- [RStudio's R Cheatsheets for Rmarkdown and Data Wrangling](#)
- [An Rmd slideshow Intro to R](#)

Open Source software and tools

- [FOSS \(Free and Open Source Software\)](#) is a copyleft approach to software which is hat is distributed in a manner that allows its users to run the software for any purpose, to redistribute copies of, and to examine, study, and modify, the source code. [12]
- [vim \(or Gvim the gui version\)](#) is a powerful text and code editor, that is universally available on all linux and mac computers.[13] [NeoVim](#) is a new Gvim fork.[14] It can be installed on windows computers, its available on the ODS Desktops.. [13]
- [Git \(Wikipedia\)](#) is a distributed content versioning system that is very popular. It enables collaborative code development and LaTeX writing projects.[15]

- [Git server software](#) is installed on each computer.[16]
- [GitHub](#) is a Git server website used for collaborative code development.[17]
- [BitBucket](#) is a Git server website used for collaborative code development. If you join with your case.edu email address, you get unlimited private repositories.[18]
- [Stack Exchange](#) [19] Code Question and Answer Websites: covering R, Python, Mathematica, LaTeX and many other things, such as English or Spanish etc.

Python3 (is also used for Data Science in many cases. But here we will focus on R first.

- [Wikipedia](#): Python is a widely used general-purpose, high-level programming language. [20]
- [The Python main website](#). [21]
- [The Python Tutorial — Python v3.8.5 documentation](#) [20]
- [The Hitchhikers Guide to Python](#). This is an open access book being hosted on developed on GitHub and is located here <https://github.com/vuylab/python-guide>. [22] [23]
- [NumPy](#) is the fundamental package [24] for scientific computing with Python.
- [FiPy](#): Partial Differential Equations with Python [25]
- [SciPy](#) is a python-based ecosystem [26] of open-source software for mathematics, science, and engineering.
- [IPython Shell and Notebook](#) [27]
- [Spyder](#) is the Scientific Python Development Environment [28]

LaTeX is used for publication quality writing. Its also the backend for Rmarkdown's pdf generation. It lets you write professional looking papers, theses and books, along with presentations.

- [LaTeX](#) is a program for writing documents, paper, journal articles, presentations and theses. [29]
- [LaTeX - Wikibooks](#), open books for an open world. [30]
- [Zotero Reference-Citation Manager, BibTeX Client](#) [31]

9 Policies

9.1 Attendance

You attendance is expected. Some information is covered that is not in the text. Student participation is an important part of the class.

9.2 Readings

Readings must be done, BEFORE the class, where they are assigned. The reading assignment, is for the class with which it is listed.

9.3 Lab Exercise Assignments

Lab Exercises are due before class on the Tuesday or Thursday they are due. A 50% deduction will be assessed for submissions not received on canvas by the due date.

9.4 Collaboration and Citation

Discussions and working together (except on exams) is acceptable and encouraged. It is not ethical to do someone else's work or to have someone do your work. You must cite all resources you used to work on your Lab Exercises and SemProjects. Citations should be done at the end of the document. These can be to books, Wikipedia and other web resources, and discussions with other students.

9.5 Academic Integrity Policy

All students in this course are expected to adhere to University standards of academic integrity. Cheating, plagiarism, misrepresentation, and other forms of academic dishonesty will not be tolerated. This includes, but is not limited to, consulting with another person during an exam, turning in written work that was prepared by someone other than you, making minor modifications to the work of someone else and turning it in as your own, or engaging in misrepresentation in seeking a postponement or extension. Ignorance will not be accepted as an excuse. If you are not sure whether something you plan to submit would be considered either cheating or plagiarism, it is your responsibility to ask for clarification.

For complete information, please go to <https://students.case.edu/community/conduct/aiboard/policy.html>.

9.6 Disability Resources

ESS Disability Resources is committed to assisting all CWRU students with disabilities by creating opportunities to take full advantage of the University's educational, academic, and residential programs.

For further information, please go to <https://students.case.edu/academic/disability/>.

10 Copyleft, References, Citations & Rubrics

10.1 CopyLeft

Creative Commons plays an important role in openness and open science, open data, open source efforts.

This DSCI351-451 class [32] is covered by a Creative Commons [33] copyleft licenses.

The license we'll use for class materials, code and presentations is covered by the "Attribution-ShareAlike 4.0 International" license, which is commonly called the CC BY-SA 4.0 license. [34]

More information on licensing open works, can be found on Wikipedia. [35]

GNU [36] is the developer of the [GPL License](#) [37] that is used for many open source software projects, such as Linux.

11 Setting up your R data science computer

To setup and install the software for an R Open Data Science workstation, here's how.

11.1 R Standard Packages for Win,Lin,Mac

This is our list of R standard packages that you will install whether you are using Windows, Linux or Macintosh Operating Systems.

- Here is the list of standard packages that we suggest you install. If a warning comes up asking whether you want to install packages from the source, answer "y" for yes.

- Alphabetical packages:

```
acepack adabag akima Amelia AmesHousing animation anytime arrow arsenal arules askpass aspace
astsa available babynames backports bagRboostR baseline BayesFactor bayesSurv BayHaz bcpa bda
BiocManager birk bit64 blavaan blogdown BMA bookdown bookdownplus BoomSpikeSlab boot bootstrap
bomrang breakDown breakpoint brickr brms broom bsts C50 Cairo callr calendR car carData
caret cartogram caretEnsemble CausalImpact causalTree centiserve cgwtools changepoint
changepoint.np ChemoSpec circlize Ckmeans.1d.dp class ClimClass cloudml cluster.datasets coda colormap
colortools CORElearn corgram corplot cowplot cowsay crayon cpc4 cranlogs ctv CVXR datapasta
dataRetrieval data.table data.tree dataMaid DBI DBITest dbplyr dbSCAN ddiv dendexten DescTools
devtools DiagrammeR DiagrammeRsvg dice dichromat digest directlabels dlstats DMwR doFuture
doMC doParallel dplyr drake drumr lidR DT dummies dtwclust e1071 easyNCDF Ecdat eemR
effects ElemStatLearn epiDisplay equatiomatic exifr extrafont fabletools factoextra FAIRmaterials
fastcluster fastDummies feasts feather flexclust flextable flipbookr forecast foreach formattable
fpp2 fpp3 fmsb fun fuzzyjoin gam gameofthrones gapminder gbm gclus geojson geojsonio geomnet
geoR geoRglm GGally ganimate ggalluvial ggbio ggbump ggcorrplot ggdag ggdark ggdendro ggdist
ggeffects ggExtra gggenealogy ggforce ggfortify gghighlight ggplot2movies ggsci ggsoccer ggstatsplot
ggsubplot ggiraph gglasso ggm ggmap ggnetwork ggnewscale ggplot2 ggpublish ggQC ggRandomForests
ggraph ggraph ggrepel ggridges ggrepel ggsankey ggseas ggsignif ggspectra ggstance ggstream ggtech
ggtern ggthemes ggTimeSeries ggtree ggvis ggvenn ggvoronoi ggdark ggtech ggalt ggraph ggpmisc
geomnet ggnetwork ggTimeSeries ggradar ggstance ggbeeswarm ggridges glmnet glmnetUtils gmodels
googlesheets4 googleVis GPArotation graphlayouts gridBase gridExtra gridGraphics gsl gstat gt
gtrendsR gutenbergr gWidgets2 HadoopStreaming HarmonicRegression hashFunction hcp hdf5r hdPCA
heatmaply here hexbin highcharter HH Hmisc hrbrthemes htr htmlwidgets huxtable
hyperSpec ICON igraph igraphdata igraphshiny imager infer ipred InformationValue IQCC IRdisplay
IRkernel ISLR ISLR2 itertools janitor jsonlite jsonvalidate jtools kableExtra keras kerasformula
kerasR kernlab keyring kgc klaR knitr citations knitr Lahman lars lavaan lavaan.survey leaps learningr
learnr learnrbook learNN learnBayes lime lintr lme4 lmerTest lobstr logitnorm magick magickGUI
magrittr Make mapdata Mapmate mapproj maps markdown maptools mapSpain mapview MASS MatrixMatrix
MatrixModels matrixStats markovchain mcmc MCMCglmm meltr metRology meme Metrics mgcv
mice minpack.lm mixtools mlbench MLmetrics mlr mnormt mosaicData MTS multiway multicolor
naniar navdata NbClust ncdf4 netSEM networkD3 neural neuralnet NeuralNetTools nFactors NLP
NMF nnet nycflights13 odbc OIdata olrr OIsurv onehot OneR onlineCPD openair openintro OpenMX
OpenStreetMap operator.tools optimx optiRum ordinal osmdata OSMscale OutlierDetection oysteR
ozmaps packageRank packHV packrat palmerpenguins parallelSVM packcircles pastecs patchwork
pca3d PCAmixdata performance PerformanceAnalytics pins pipeR pixiedust paletteer plot3D
plotmo plotKML plotly plotrix plotROC pls plsdepot Plumber plyr plyrMr png pool pracma prcomp
prodlim pROC processx prophet profvis propagate proxy pryr psych purrr PVplr pwr qcc qqplotr
qtlmt qualityTools quantmod r2d3 ragg randomcoloR randomForest randomForestExplainer randomForestSRC
randomNames ranger raster rasterVis rayshader rCharts RColorBrewer Rcpp RCurl Rdice
Rdpack readr recipes RefManageR regclass relaimpo remotes repr reshape reshape2 reactable
reticulate revealjs Rfast rgdal rgeos rggobi rgl RgoogleMaps rhub rJava rjson RJSONIO rlang rlist
```

RLRsim rmapshaper rmarkdown Rmisc Rmpi rms RMySQL RNiftyReg RNHANES rNMF roxygen2 rockchalk ROCR rpart rpart.plot rpf rprojroot rPython rsample RSNNS rstan rstanarm rsvg RTest RTextTools rticles rTorch Rtsne rtweet RUnit rvest rworldmap rworldxtra scatterplot3d scico scrypt seasonal segmented sem semPlot semPLS sf showtext shiny shinydashboard shinyjs shinystan shinytest shinythemes signal simpleNeural SixSigma sjstats skimr slackr sloop sp spacetime spacyr sparklyr sparkline sparktf spc spectacles spelling sqldf sqliter sqldfutils staRdom stargazer stars stationaRy statsr stlplus stockPortfolio stopwords streamgraph StreamMetabolism stringi stringr styler suncalc SunsVoc sunburstR survival survivAll survivalAnalysis survivalMPL survivalROC survivalsvm survminer svglite svUnit SwarmSVM synthpop targets TeachingDemos TeachingSampling tensorflow testthat textdata tfdatasets tfdeploy tfestimators tfruns tibble tictoc tidygraph tidymodels tidyposterior tidyR tidytext tidyverse tidyverts timeDate timeevs tinytex tipr tm torch torchdatasets torchvision transformr tree treemap treemapify tsibble TSclust TSstudio tweeNR usmap V8 validate vcd vioplot viridis visNetwork vtreat waterfalls WaveletComp wavelets wavethresh weathermetrics webshot withr wmtsa WGCNA WDI wordcloud xaringan xkcd xgboost XLConnect XML xtable xts yardstick zeallot zipcode zoo

You can go the the highlighted tab in below picture and install/upgrade you packages here. To install, simply paste the list of packages in the window.

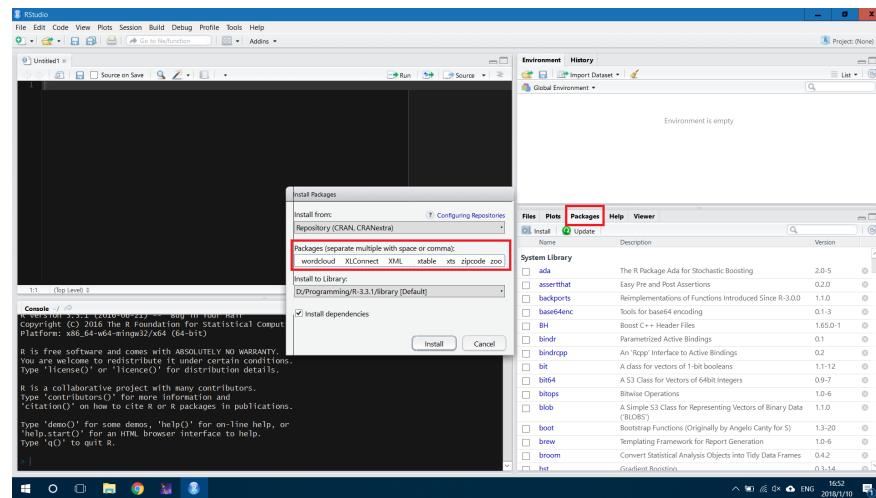


Figure 10

To upgrade your packages, select all packages and press "Install Updates"

Note that the "R standard packages are listed under the "For Windows" section in subsubsection 11.1.

11.2 For Windows

In Windows we are allowed to use spaces in filenames, however, most other systems does not support that. To avoid conflicts or troubles, we suggest using **camelBack** naming convention or use "-" or "_" to replace spaces.

11.2.1 LaTeX

L^AT_EX is a document preparation system that is widely used in the academia for producing scientific documents. You will need to install two softwares, Miktex and TeXstudio.

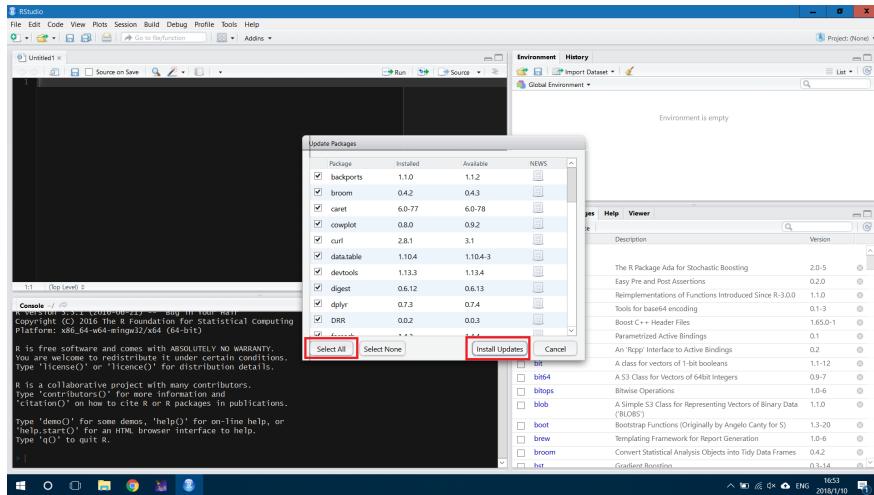


Figure 11

- Download and run the Basic MiKTeX Installer. MiKTeX has the ability to install missing packages automatically, i.e., this installer is suitable for computers connected to the Internet. Before you run the installer, you can check the [prerequisites](#). The installer is available on the [download](#) page. You start it with a double-click on the downloaded file.
- Read the Copying Conditions carefully and click "I accept the MiKTeX copying conditions", the click "Next", as demonstrated below.

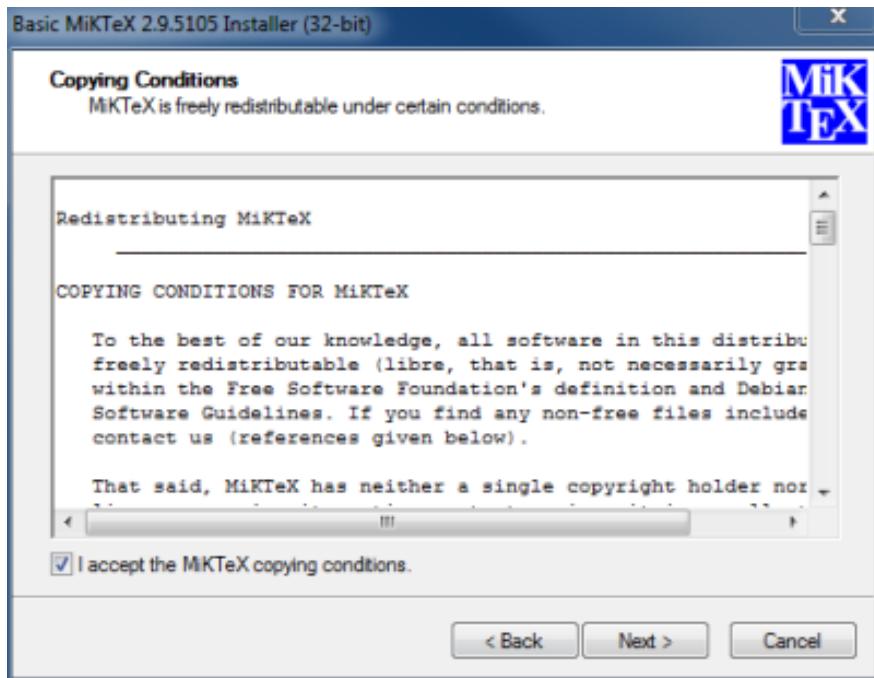


Figure 12

- You have the Option to create a shared MiKTeX installation. Click "Anyone who uses this Computer (all users)", if you want to install MiKTeX for all users. Click "Only for ...", if you want to install MiKTeX for yourself only. When you have made your decision, click "Next" to go to the next page.

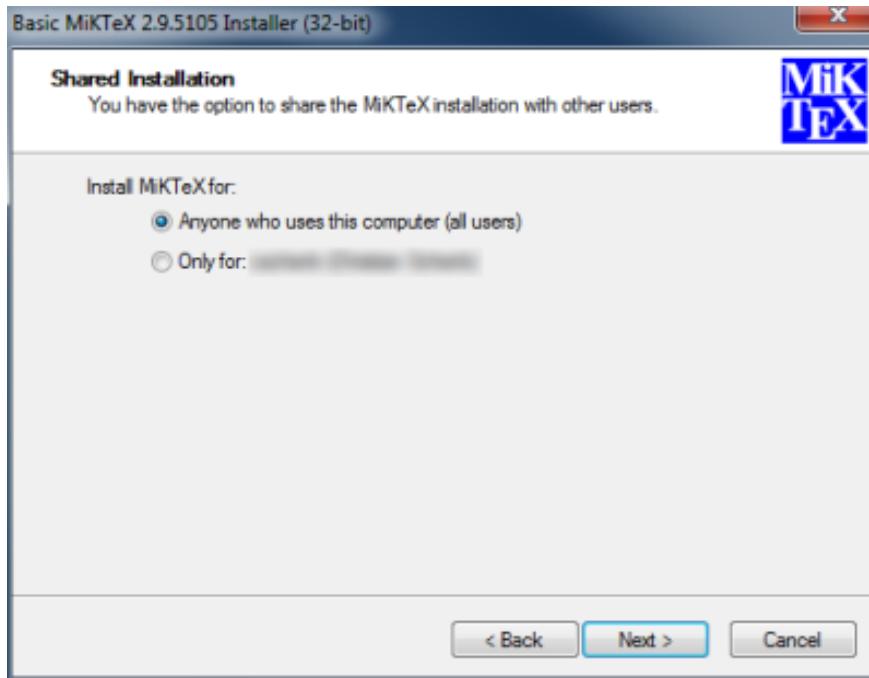


Figure 13

- You can specify the directory where you want to install your Miktex. Click "Browse", if you want to specify another (than the default) directory location. Click "Next", to go to the next page.

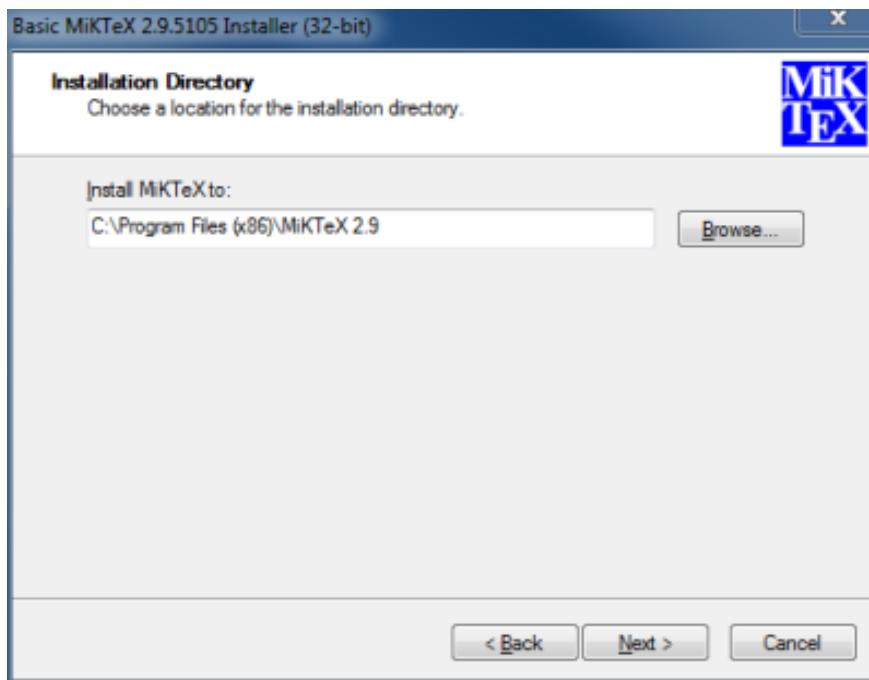


Figure 14

- The installer allows you to set the preferred paper size (usually it's A4 in China and letter size in the US). You also have the option to change the default behavior of the integrated package manager for the case where a required package is missing. Select "Yes", to make the package manager is always allowed to install missing packages. All these configurations can be changed later.

Click "Next", to go to the next page.

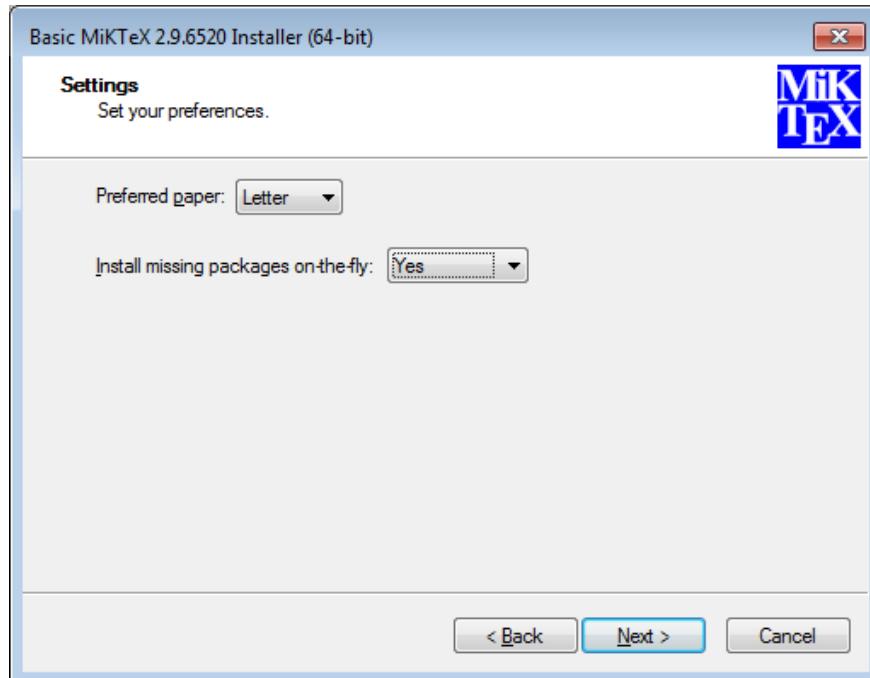


Figure 15

- Before the actual installation process begins, you get a chance to review your decisions. If you are satisfied with the settings, then click "Start" to start the actual installation.
- The installation will take a few minutes. The progress bar shows an approximate percentage of completion. When the installation has finished, you can click "Next" to open the last page.
- MiKTeX is now installed. Click "Close", to close the installer.
- In order to make use of latex the easiest way is to use a integrated development editor (IDE). **TeXstudio** is a free package that allows you to edit tex documents, compile and view them, it has syntax highlighting, auto completion, in line spell and grammar checker and much more. You can find the downloads page [here](#) and click on **download now**.
- Once downloaded, run and start the installer.
- Accept all the default conditions, and start up TeXstudio to finish.
- If you need instructions on how to start using LaTeX, here are some [tutorials](#).

11.2.2 Git

Git Bash is command line programs which allow you to interface with the underlying git program. Bash is a Linux-based command line, which has been ported over to Windows.

- Download latest version of Git Bash on the [official website](#).
- Once Git Bash Windows installer is downloaded, run the executable file and follow the setups:
- Agree to the GNU General Public License and click "Next".



Figure 16

- Select the components you want to install and click Next. We suggest that you should unselect Windows Explorer integration.

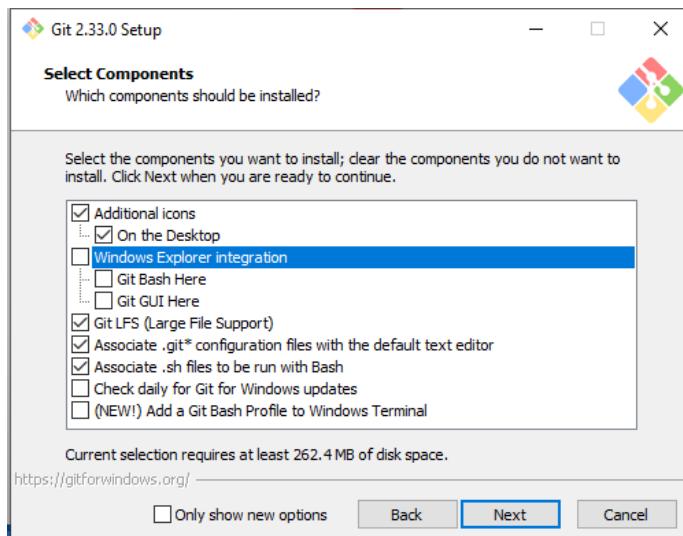


Figure 17

- Set default editor to Vim(which is the default option).

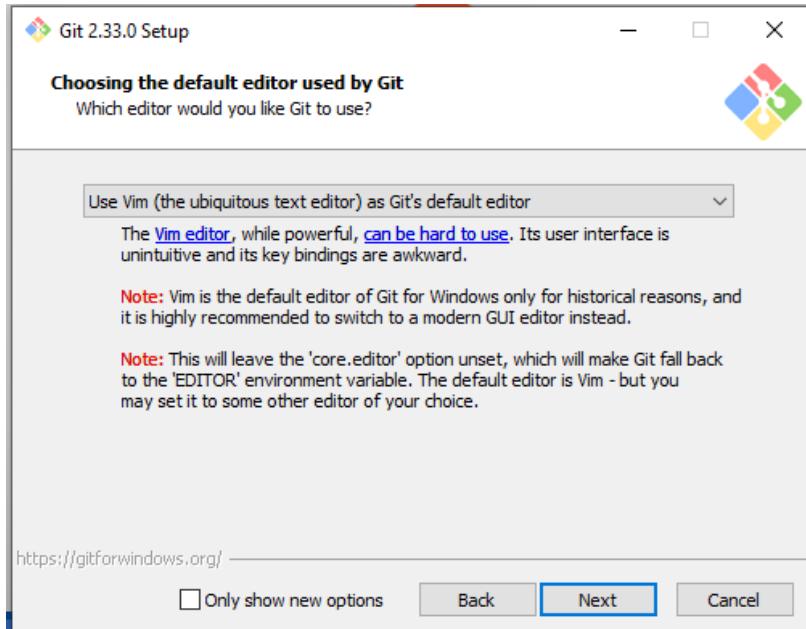


Figure 18

- Choose default initial branch name.

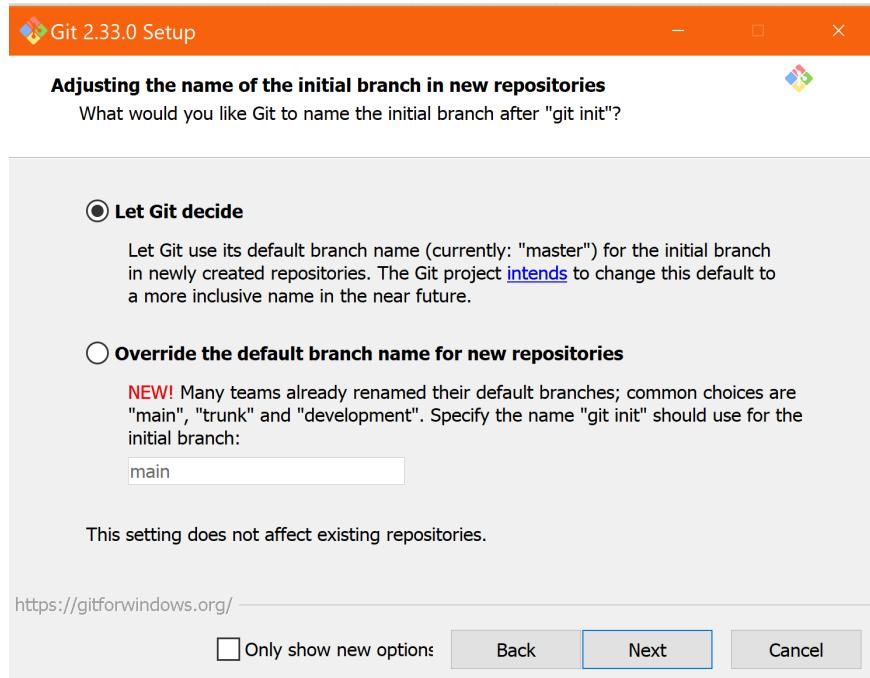


Figure 19

- We suggest that you use the default option, which is "Use Git from Git Bash only".

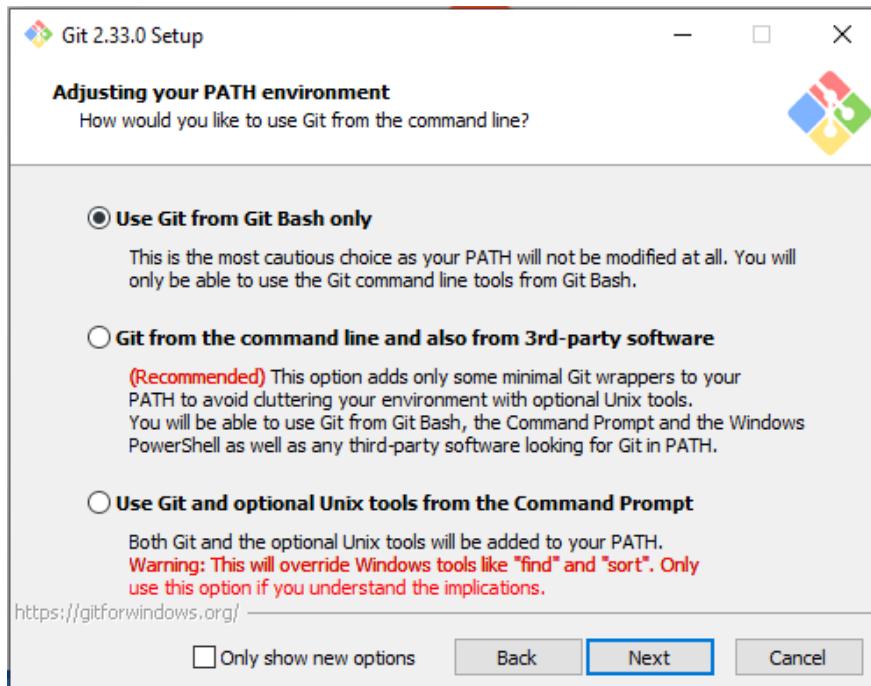


Figure 20

- Select which SSH client program to use.

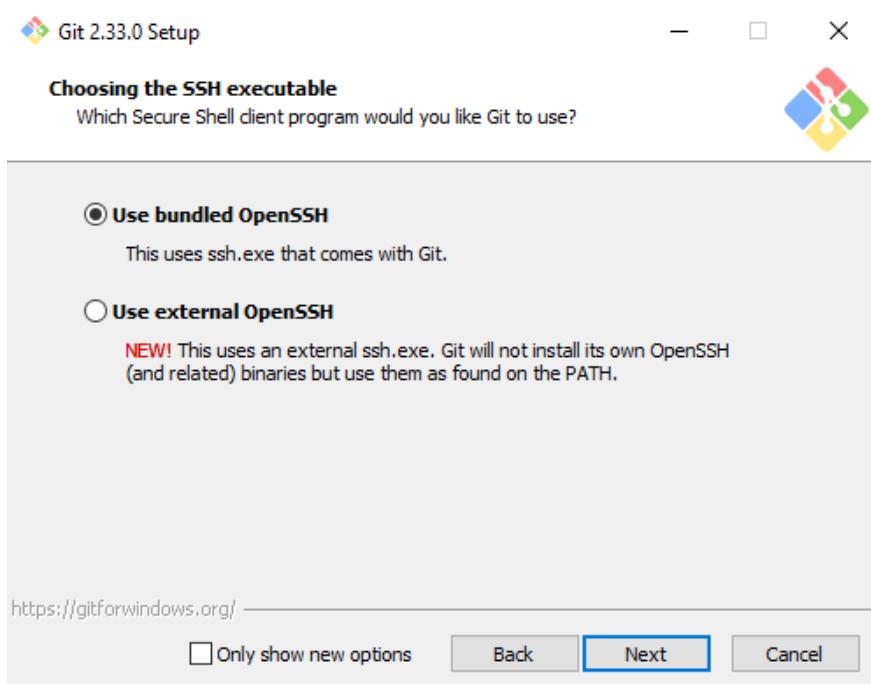


Figure 21

- Select which SSL/TLS library would you like to use for HTTPS connection and click Next.

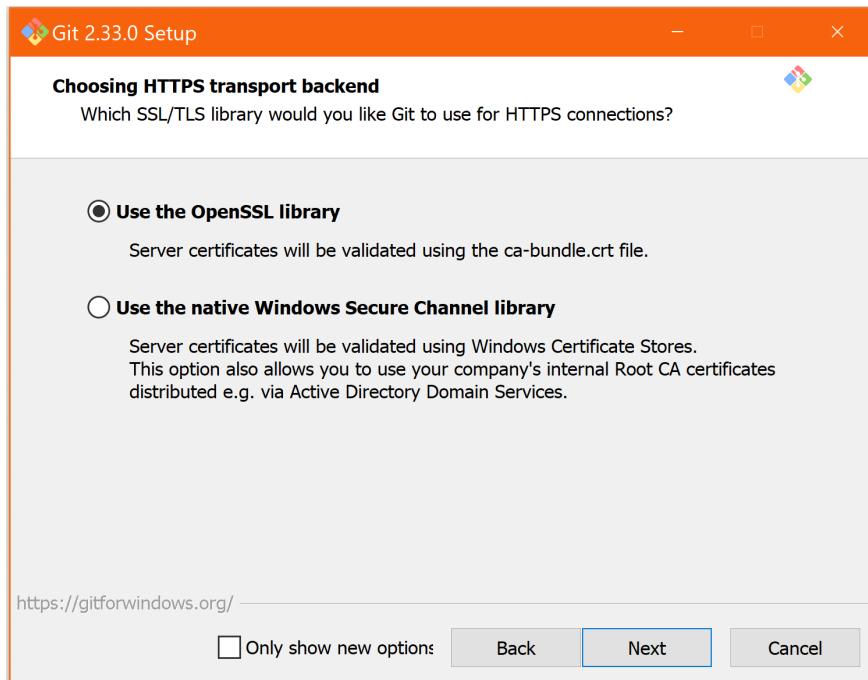


Figure 22

- Select, how should Git treat line endings in text files and click Next.

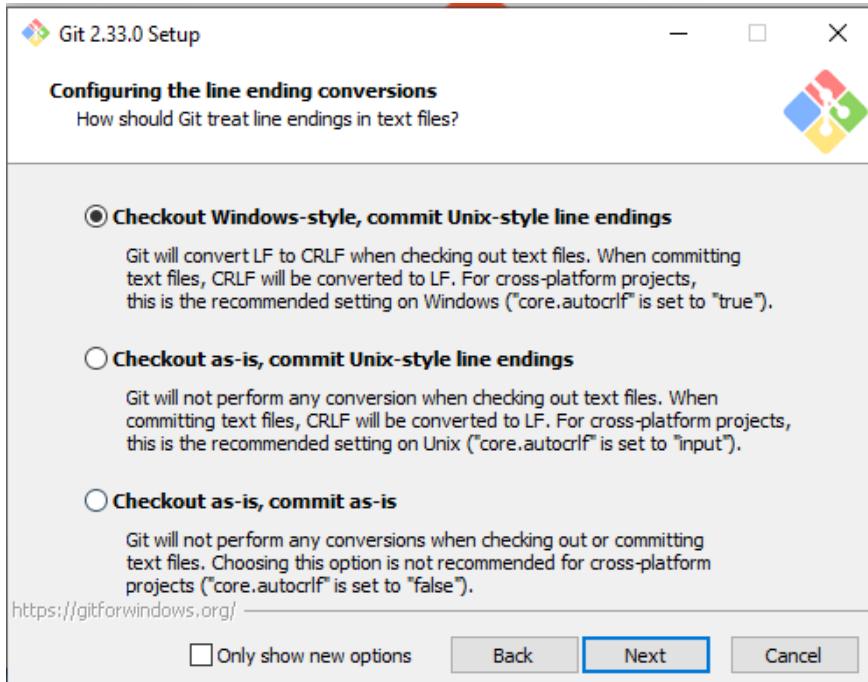


Figure 23

- Select the terminal you want to use for Git Bash.

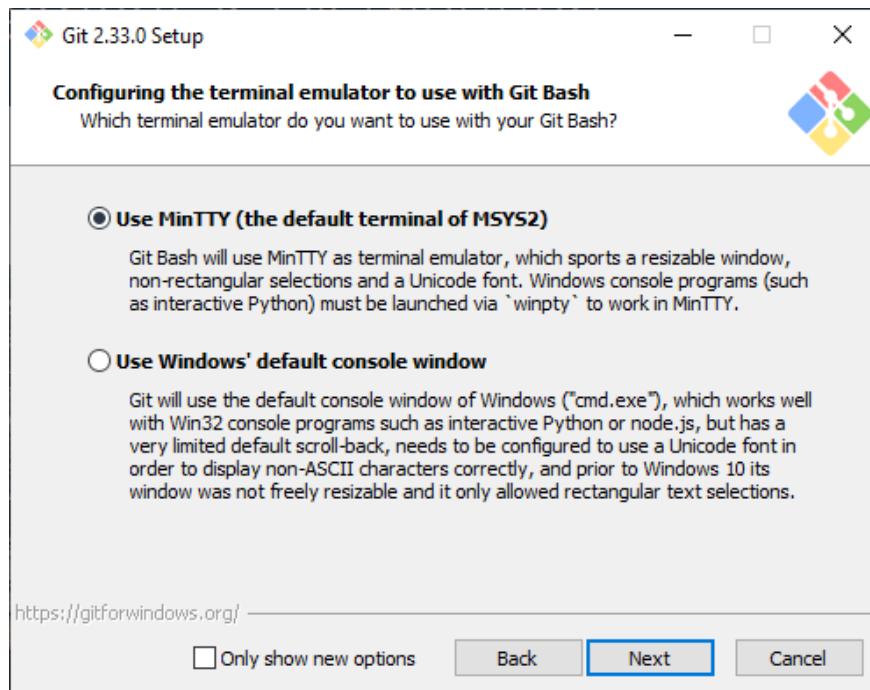


Figure 24

- Select the default behavior of 'git pull'

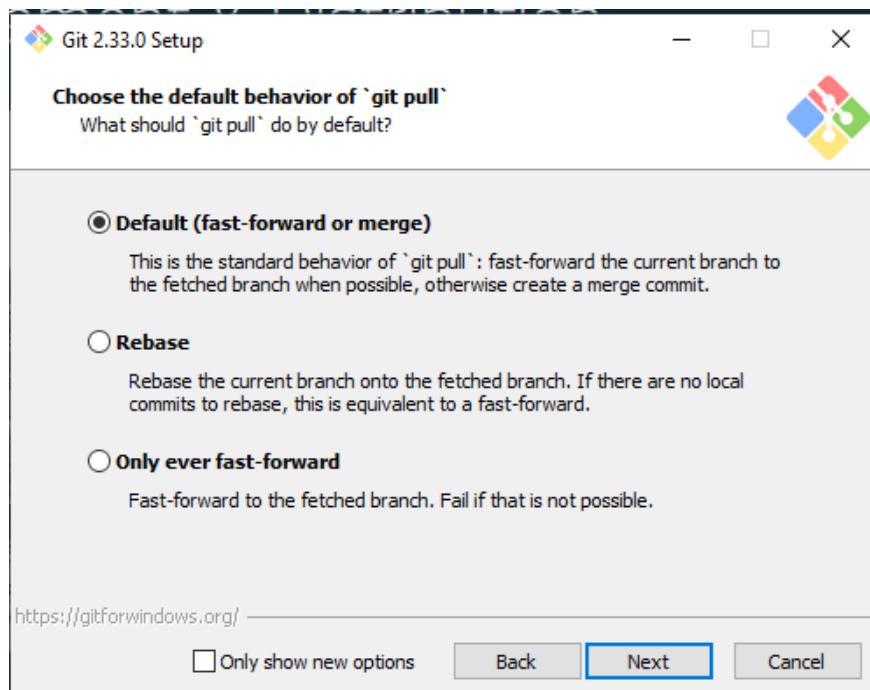


Figure 25

- Select the features you want to enable and click "Next". We suggest that you unselect "Enable Git Credential Manager".

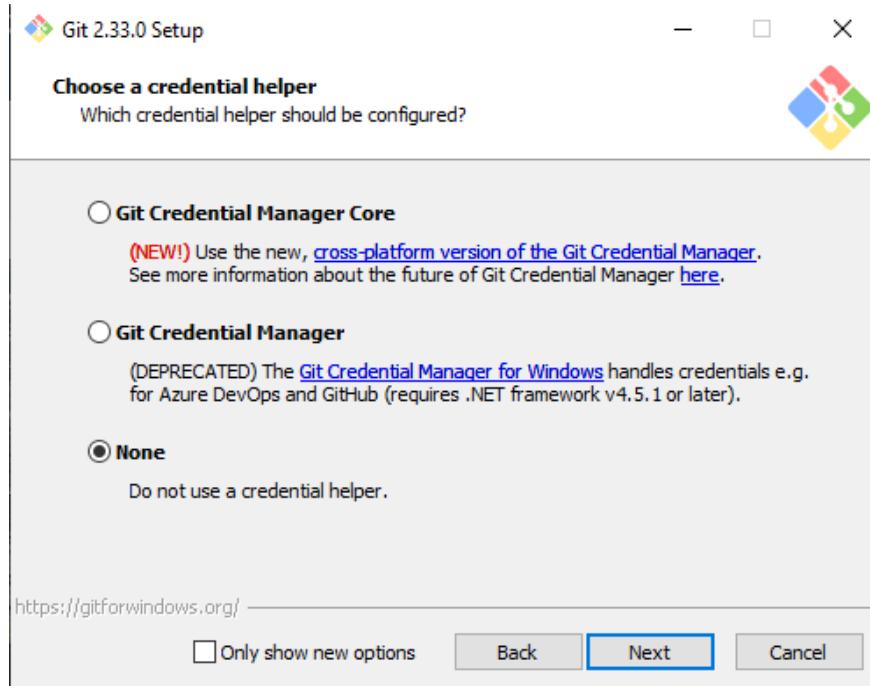


Figure 26

- Unselect the experimental options.

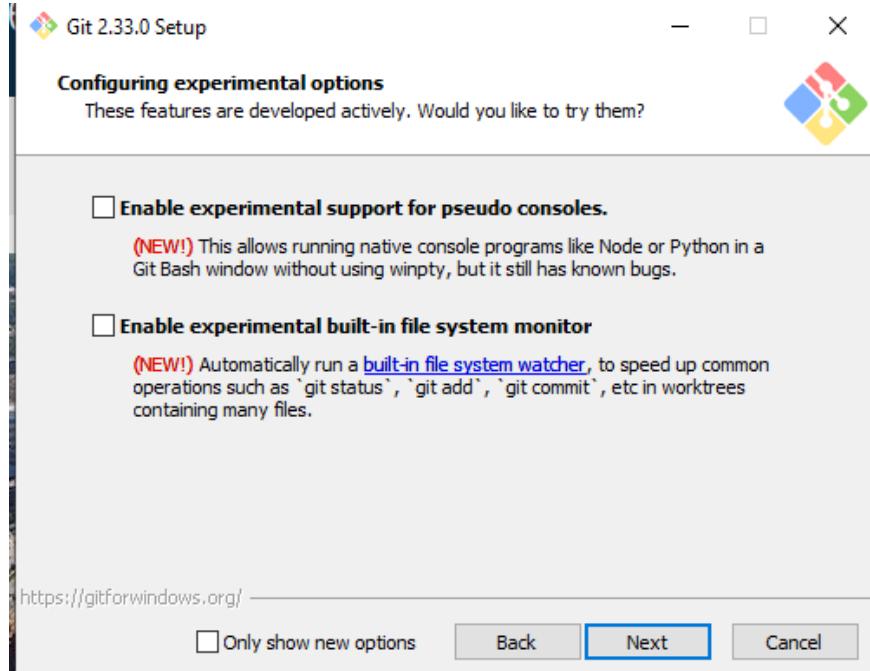


Figure 27

- Please wait while Setup wizard installs Git on your computer and click "Finish" to exit the Setup wizard.
- After Git Bash installation finishes you will ready to use the Linux command on a windows machine. Double click on below icon to start the Git Bash.

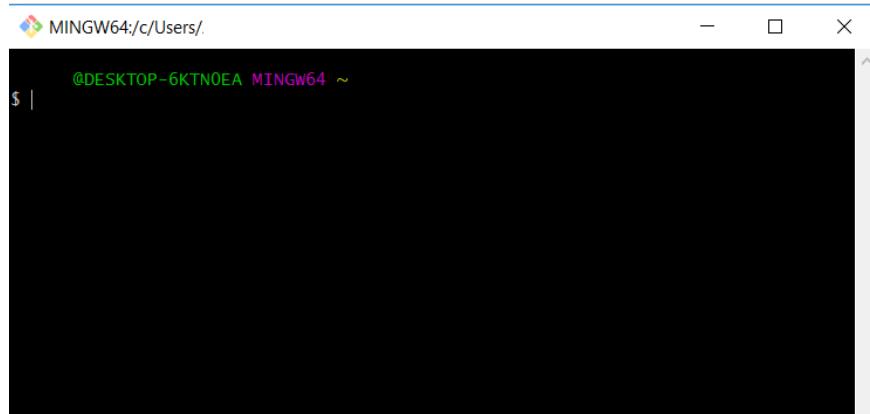


Figure 28

- Set up user name and email in Git.

```
git config --global user.name "yourusername"
git config --global user.email "youremail@website.com"
git config --global color.ui auto
```

- Here are some common commands you can use in git:

pwd – present working directory
 cd – change directory
 ls – list files in current working directory
 mkdir – make new directory

- If you want to learn more about how git works (pull request, merge and more), you can read some [tutorials](#).

11.2.3 R

R is a free programming language and software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing, while RStudio is a free and open-source integrated development environment for R.

- To [download R](#), please choose your "install R for Windows" and then choose base R for a complete installation.
- Double click on the installer, and follow the instructions.
- Users of Vista/Windows 7/8/Server 2008/2012 installing for a single user using an account with administrator rights should consider installing into a non-system area (such as C:\R).
- Please try to avoid spaces or any special characters other than English letters and numbers in your installation directory, which may cause error later.
- After installing R, you can download [Rstudio here](#), and choose the RStudio Desktop Open Source License version (the left most one).

- Run the installer and follow the installation instructions.
- Again, please try to avoid spaces or any special characters other than English letters and numbers in your installation directory.
- Rstudio have some built-in packages such as tidyverse and ggplot2, but if you are interested in building your own R packages, you can [download Rtools](#). Please choose the latest version, as the older versions are not compatible with latest release of R.
- Run the installer, and accept the defaults throughout.
- Confirm and finish the installation.
- Once the Rtools installation completes, open RStudio and go to Profile–Global options–Code and change the code editing options as follows:

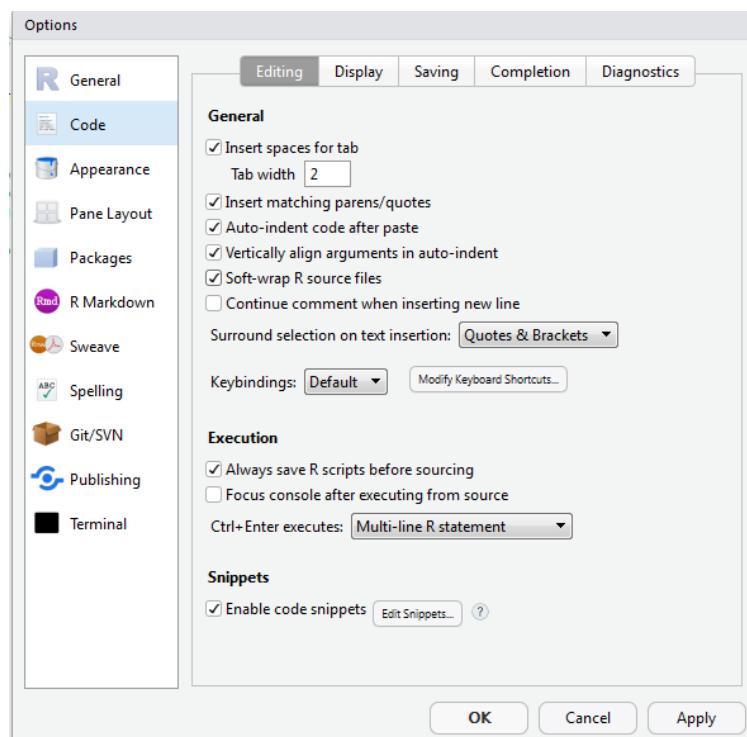


Figure 29

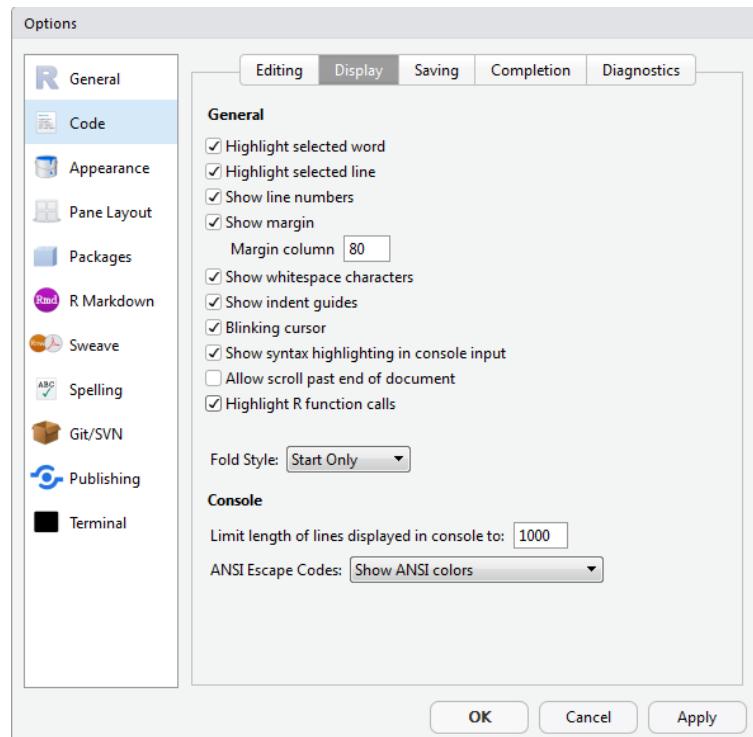


Figure 30

- You can also change your appearance style in Global Options:

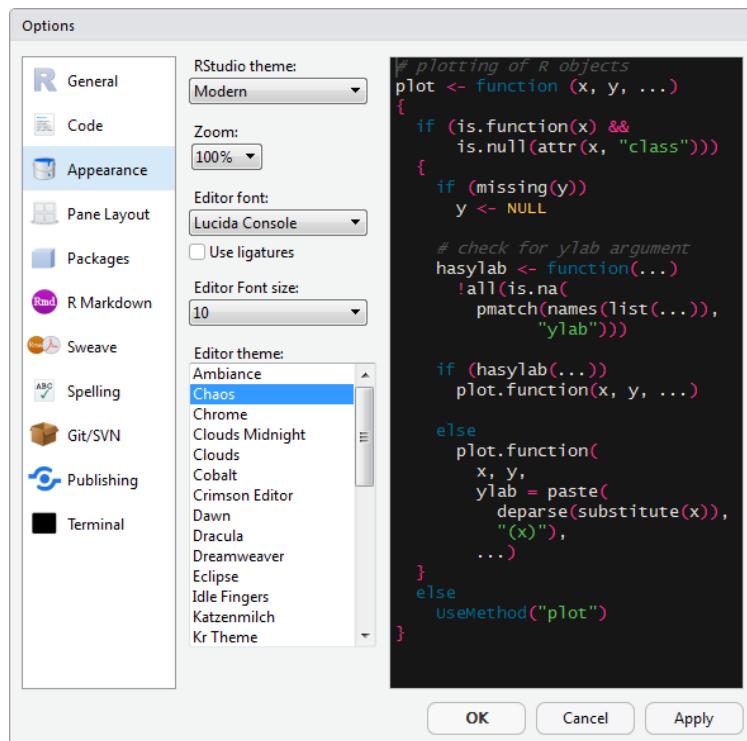


Figure 31

11.2.4 GVim

GVim offers a graphic user interface for the editor **Vim**. This is a powerful editor but could be a little bit hard to use.

- You can download Vim from their [download page](#). For Windows system, click on "PC: MS-DOS and MS-Windows", and download "gvim80.exe".
- Open the installer and accept the default conditions.

11.3 For Linux

11.3.1 LaTeX

TeX Live is an easy way to get up and running with the TeX document production system, it is available on most Unix-like systems, but it is recommended to use MacTeX if you are using Mac OSX. To install TeXLive and TeXstudio, run the following code:

```
sudo apt-get install texlive-full texworks texstudio
```

11.3.2 Git

To install Git, run the following code:

```
sudo apt-get install git
```

11.3.3 R

- Install from CRAN:

```
## This sets up the CRAN repository in your Linux Package Manager
sudo echo "deb http://cran.rstudio.com/bin/linux/ubuntu xenial/" |
sudo tee -a /etc/apt/sources.list
gpg --keyserver keyserver.ubuntu.com --recv-key E084DAB9
gpg -a --export E084DAB9 | sudo apt-key add -
sudo apt-get update
sudo apt-get install r-base r-base-dev
## extra linux packages needed by
sudo apt-get install r-cran-xml pkg-config libxml2-dev
libtiff5-dev fftw3 fftw3-dev tmux
cifs-utils openssh-server openssh-client tree htop
gdebi curl libcurl4-dev openssl-dev libssl-dev ffmpeg
```

```
sdle@vuv54:
$ gpg --keyserver keyserver.ubuntu.com --recv-key E084DAB9
gpg: key 4359ED62E084DAB9: public key "Totally Legit Signing Key <mall
ory@example.org>" imported
gpg: Total number processed: 1
gpg:           imported: 1
sdle@vuv54:
$ gpg -a --export E084DAB9 | sudo apt-key add -
OK
sdle@vuv54:
$ sudo apt-get update
Get:2 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran35/ InRe
lease [3,626 B]
Hit:3 http://us.archive.ubuntu.com/ubuntu bionic InRelease
Hit:4 http://archive.canonical.com/ubuntu bionic InRelease
Get:5 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.
7 kB]
Ign:6 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1
804/x86_64 InRelease
Get:7 http://cran.rstudio.com/bin/linux/ubuntu xenial/ InRelease [3,60
7 B]
```

Figure 32

```
sdle@vuv54:
$ sudo apt-get install r-base r-base-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
r-base-dev is already the newest version (3.6.3-1bionic).
r-base-dev set to manually installed.
r-base is already the newest version (3.6.3-1biono). →
0 upgraded, 0 newly installed, 0 to remove and 2 not upgraded.
```

Figure 33

```
sdle@vuv54$ sudo apt-get install r-cran-xml pkg-config libxml2-dev libt
iff5-dev fftw3 fftw3-dev tmux cifs-utils openssh-server openssh-client
tree htop gdebi curl libcurl4-openssl-dev libssl-dev ffmpeg
Reading package lists... Done
Building dependency tree
Reading state information... Done
Note, selecting 'libfftw3-3' instead of 'fftw3'
Note, selecting 'libfftw3-dev' instead of 'fftw3-dev'
htop is already the newest version (2.1.0-3).
libfftw3-dev is already the newest version (3.3.7-1).
pkg-config is already the newest version (0.29.1-0ubuntu2).
gdebi is already the newest version (0.9.5.7+nmu2).
libfftw3-3 is already the newest version (3.3.7-1).
tree is already the newest version (1.7.0-5).
cifs-utils is already the newest version (2:6.8-1ubuntu1.1).
curl is already the newest version (7.58.0-2ubuntu3.14).
libcurl4-openssl-dev is already the newest version (7.58.0-2ubu→tu3.14
).
libssl-dev is already the newest version (1.1.1-1ubuntu2.1~18.04.10).
libtiff5-dev is already the newest version (4.0.9-5ubuntu0.4).
```

Figure 34

- Before installing, you should [check the latest version](#) of RStudio, and change the version number in the code below accordingly. Install RStudio:

```
## Update to the latest version number in the lines below
wget https://download1.rstudio.org/rstudio-1.4.1717-amd64.deb
sudo gdebi -n rstudio-1.4.1717-amd64.deb
rm rstudio-1.4.1717-amd64.deb
```

The screenshot shows a terminal session on a Linux system. The user runs `cd home/sdle/Downloads/` and then `sudo apt-get install gdebi-core`. They then use `wget` to download the RStudio server package from the official website. The package is 57613138 bytes (55M) in size and is saved as `rstudio-server-1.4.1717-amd64.deb`. The download completes in 1.4 seconds at 38.6MB/s.

```
sdl@vuv54:~$ cd home/sdle/Downloads/
sdl@vuv54:~/Downloads$ sudo apt-get install gdebi-core
Reading package lists... Done
Building dependency tree
Reading state information... Done
gdebi-core is already the newest version (0.9.5.7+nmu2).
0 upgraded, 0 newly installed, 0 to remove and 2 not upgraded.
sdl@vuv54:~/Downloads$ wget https://download2.rstudio.org/server/bionic/amd64/rstudio-server-1.4.1717-amd64.deb
--2021-08-17 15:55:23-- https://download2.rstudio.org/server/bionic/amd64/rstudio-server-1.4.1717-amd64.deb
Resolving download2.rstudio.org (download2.rstudio.org)... 99.84.189.44, 99.84.189.81, 99.84.189.124, ...
Connecting to download2.rstudio.org (download2.rstudio.org)|99.84.189.44|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 57613138 (55M) [application/x-deb]
Saving to: 'rstudio-server-1.4.1717-amd64.deb'

rstudio-server-1.4.1717-amd64.de 100%[=====] 54.94M 38.6MB/s   in 1.4s
2021-08-17 15:55:24 (38.6 MB/s) - 'rstudio-server-1.4.1717-amd64.deb' saved [57613138/57613138]
```

Figure 35

The screenshot shows a terminal session where the user runs `sudo gdebi rstudio-server-1.4.1717-amd64.deb`. The process installs the RStudio server package, which includes the RStudio application and its dependencies. The server starts successfully, and its status is shown as active (running).

```
sdl@vuv54:~$ sudo gdebi rstudio-server-1.4.1717-amd64.deb
Reading package lists... Done
Building dependency tree
Reading state information... Done
Reading state information... Done

RStudio Server
RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, and workspace management.
Do you want to install the software package? [y/N]:y
Selecting previously unselected package rstudio-server.
Preparing to unpack rstudio-server-1.4.1717-amd64.deb ...
Unpacking rstudio-server (1.4.1717) ...
Setting up rstudio-server (1.4.1717) ...
Created symlink /etc/systemd/system/multi-user.target.wants/rstudio-server.service → /lib/systemd/system/rstudio-server.service.
● rstudio-server.service - RStudio Server
   Loaded: loaded (/lib/systemd/system/rstudio-server.service; enabled; vendor preset: enabled)
   Active: active (running) since Tue 2021-08-17 15:55:50 EDT; 1s ago
     Process: 27949 ExecStart=/usr/lib/rstudio-server/bin/rserver (code=exited, status=0/SUCCESS)
    Main PID: 27950 (rserver)
      Tasks: 3 (limit: 4915)
     CGroup: /system.slice/rstudio-server.service
             └─27950 /usr/lib/rstudio-server/bin/rserver
                 ├─27950 /usr/lib/rstudio-server/bin/rserver
                 ├─27954 /bin/sh -c /usr/bin/R --slave --vanilla -e "cat(paste(R.home('home'),R.home('share'),R.home('include'),R.home('...')))
                 └─27955 /usr/lib/R/bin/exec/R --slave --vanilla -e cat(paste(R.home('home'),R.home('share'),R.home('include'),R.home('...')))

Aug 17 15:55:50 vuv54 systemd[1]: Starting RStudio Server...
Aug 17 15:55:50 vuv54 systemd[1]: Started RStudio Server.
```

Figure 36

- Once Rstudio is installed, Install the “Standard R Packages” listed here in Section [11.1](#).

11.4 For Mac

11.4.1 Homebrew

Homebrew is a package manager for Mac OS. To install Homebrew, paste and run the following command in terminal:

```
/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

You can read more about Homebrew [here](#).

11.4.2 XQuartz

To correctly set up your linux environment, you should also install XQuartz. XQuartz is Apple Inc.’s version of the X server, a component of the X Window System for macOS. You can [download](#) and install the latest version of XQuartz.

11.4.3 LaTeX

To install LaTeX on Mac, you need to install MacTeX and TeXstudio.

- The current distribution as of today (November 29, 2022) is MacTeX-2017. This distribution requires Mac OS 10.10, Yosemite, or higher and runs on Intel processors. To download, click [MacTeX Download](#).
- After downloading, double click on the MacTeX.pkg to install. Follow the straightforward instructions. Installation on a recent Macintosh takes four to six minutes.
- At the end of installation, the installer will report "Success." But sometimes, the installer puts up a dialog saying "Verifying..." and then the install hangs. In all cases known to us, rebooting the Macintosh fixes this problem. After the reboot, install again.
- Now you can start installing TeXstudio. You can find the corresponding installer on the [TeXstudio website](#).
- Because the developers of TeXstudio do not have an Apple Developer Account, OS X may complain about an unidentified developer and deny opening TXS. In that case, open the context menu on the TXS icon (Ctrl + Click) and select open.

11.4.4 Git

There are several ways to install Git on a Mac. In fact, if you've installed XCode (or it's Command Line Tools), Git may already be installed. To find out, open a terminal and enter git –version.

Apple actually maintain and ship their own fork of Git, but it tends to lag behind mainstream Git by several major versions. You may want to install a newer version of Git using the method below:

- Download the latest Git for [Mac installer](#).
- Follow the prompts to install Git.
- Open a terminal and verify the installation was successful by typing git –version.
- Configure your Git username and email using the following commands, replacing "yourusername" with your own. These details will be associated with any commits that you create:

```
$ git config --global user.name "yourusername"
$ git config --global user.email "youremail@website.com"
```

11.4.5 R

- Download R from [CRAN](#) and click "Download R for (Mac) OS X".
- Follow the instructions and install R.
- Download the latest RStudio from their [website](#). Open the installer and follow the instructions.
- Once Rstudio is installed, Install the "Standard R Packages" listed here in Section??.

References

- [1] R. D. Peng, *R Programming for Data Science*. Leanpub, Feb. 2014.
- [2] R. D. Peng, *Exploratory Data Analysis with R*. Leanpub, Apr. 2015.
- [3] David M. Diez, Christopher D. Barr, and Mine Çetinkaya-Rundel, *OpenIntro Statistics: Third Edition*. S.l.: OpenIntro, Inc., 3 edition ed., July 2015.

- [4] H. Wickham and G. Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, 1 edition ed., Jan. 2017.
- [5] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics, New York: Springer, 1st ed. 2013, corr. 5th printing 2015 edition ed., Aug. 2013.
- [6] C. Portal, "CWRU CSE portal for VDIs and XenApps," 2014.
- [7] R, "R (programming language)," Aug. 2014. 00000 Page Version ID: 621330268.
- [8] R. Project, "The r project for statistical computing," 2014.
- [9] RStudio, "RStudio," 2014. 00000.
- [10] R. Peng, "Computing for data analysis: Week 1 - YouTube," 2014.
- [11] P. Torfs and C. Brauer, "A (very) short introduction to r," 2014.
- [12] "Portal:free software," Sept. 2014. 00000 Page Version ID: 581465934.
- [13] "Gvim online," 2014. 00000.
- [14] "Neovim," 2014. 00000.
- [15] "Git (software) - wikipedia, the free encyclopedia," 2014. 00000.
- [16] "Git," 2014. 00027.
- [17] "GitHub," 2014. 00004.
- [18] "Bitbucket: Free source code hosting for git," 2014. 00000.
- [19] S. Exchange, "Stack exchange," 2014.
- [20] T. Python, "The python tutorial — python v2.7.8 documentation," 2014. 00000.
- [21] Python, "Python.org," 2013.
- [22] "The hitchhiker's guide to python! — the hitchhiker's guide to python," 2014. 00000.
- [23] "kennethreitz/python-guide," 2014. 00000.
- [24] NumPy, "NumPy — numpy," 2014.
- [25] J. E. Guyer, D. Wheeler, and J. A. Warren, "FiPy: Partial differential equations with python," *Computing in Science & Engineering*, vol. 11, pp. 6–15, May 2009.
- [26] SciPy, "SciPy.org — SciPy.org," 2014.
- [27] IPython, "IPython shell and notebook," 2014.
- [28] Spyder, "Spyder is the scientific PYthon development environment," 2014. 00000.
- [29] "TeX users group (TUG)," 2014. 00000.
- [30] LaTeX, "LaTeX - wikibooks, open books for an open world," 2014. 00000.
- [31] Zotero, "Zotero reference/citation manager, BibTeX client," 2014.
- [32] R. H. French, "DSCI351-451: Exploratory data analysis for energy & manufacturing," 2016.
- [33] C. Commons, "Creative commons - about the licenses," 2014.
- [34] "Creative commons — attribution-ShareAlike 4.0 international — CC BY-SA 4.0," 2015.

- [35] C. Commons, “Creative commons license,” Aug. 2014. 00000 Page Version ID: 618703231.
- [36] Gnu, “gnu.org,” 2014. 00007.
- [37] G. GPL, “GNU general public license,” Aug. 2014. 00000 Page Version ID: 622300724.