

Introduction to Analytics

ALY6000 CRN 71627

Tom Breur

Week 1, Tuesday 21-SEP-2021

COVID-19 Vaccination Record Card

Please keep this record card, which includes medical information about the vaccines you have received.

Por favor, guarde esta tarjeta de registro, que incluye información médica sobre las vacunas que ha recibido.



BREUR

ANTONIUS

L

First Name

MI

Date of birth

Patient number (medical record or IIS record number)

Vaccine	Product Name/Manufacturer Lot Number	Date mm dd yy	Healthcare Professional or Clinic Site
1 st Dose COVID-19	Pfizer ER8731	04/13/21 mm dd yy	NWH Tripadvisor
2 nd Dose COVID-19	Pfizer	05/04/21 mm dd yy	NWH
Other	EW0179	____/____/____ mm dd yy	Tripadviso
Other		____/____/____ mm dd yy	

Agenda

- Introductions
- Administrative notes
- Installing R (R-Studio)
- Teaching format
- Suggestions on how to succeed
- Preparation for Week #2
 - Complete Academic Integrity quiz
 - Reading assignments
 - Discussion board: “Introduce yourself”
 - Module 1 statistical practice
 - Module 1 Executive Summary Report

Introductions: Tom Breur



$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x/\ln(x)} = 1$$
$$\sum_{k=0}^n \binom{n}{k} = 2^n$$
$$\frac{1}{c} \int f(z) dz = 2\pi i (N - P)$$



WHERE I'VE STUDIED



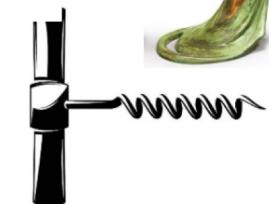
WHERE I'VE LIVED



WHERE I'VE WORKED



WHAT I DO FOR FUN



Introductions

“Networking” is an integral and valuable part of
your professional development

Welcome!

- Please take a moment to introduce yourself to your classmates on the Canvas Discussion Forum
 - https://northeastern.instructure.com/courses/95279/discussion_topics/1247260
- Building and nurturing (!) a professional network is extremely valuable for your career development
 - Note: “networking” is (a lot!) more than superficial social pleasantries at “networking events”
- Key principle is: “You have to give to get”
- Adding a picture or video helps others relate to you
 - What is unique/special about *you*?
- Where possible point to links or references that highlight your professional or personal achievements

Administrative notes

Important!!

- Due to (very!) late assignment of class section, syllabus will be subject to change
- “General” course content will *not* change, but notably **submission deadlines** of last two weeks need to be altered to enable students to succeed
- Final grades are due 11/1, and hence deadlines need to be “moved up” to accommodate this NEU deadline

Textbook Bluman

- Older versions of the textbook books may be used, and will generally be effective; however, the “connection” to page references will obviously be lost

Ed.: 3

4

5

6

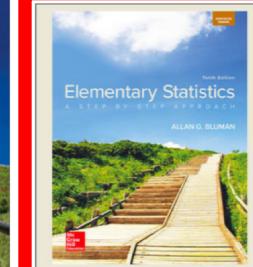
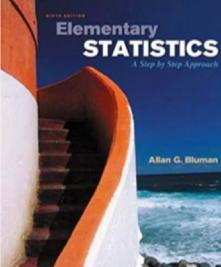
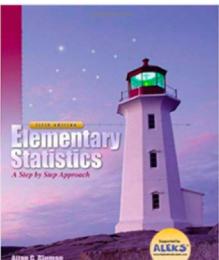
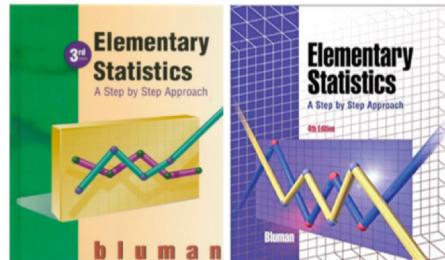
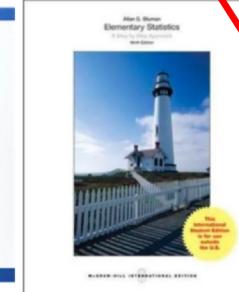
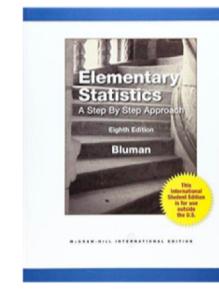
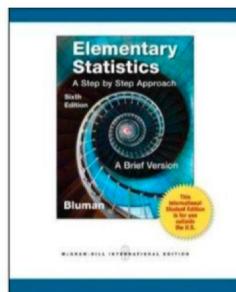
7

8

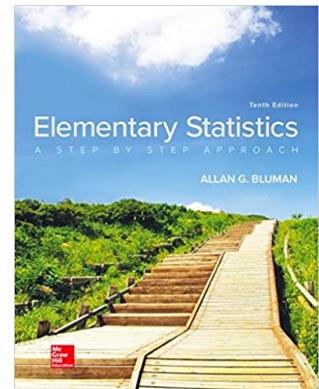
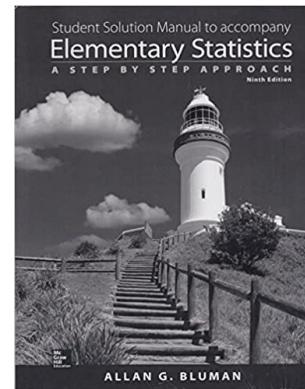
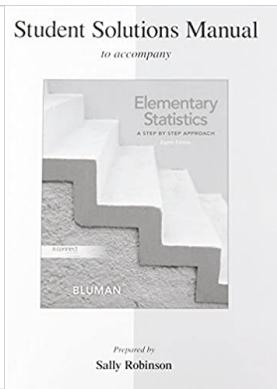
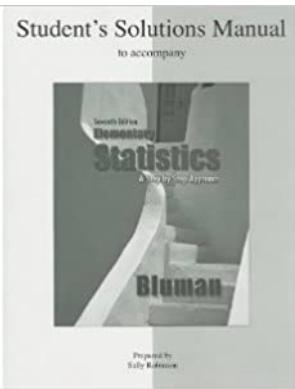
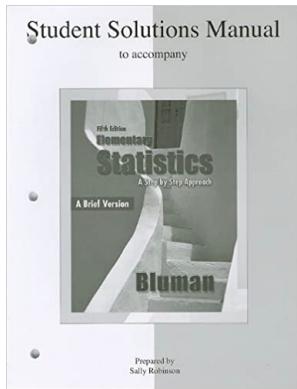
9

10

Version in use now



Student Solution Manuals



Student solution manuals only “work” with the corresponding version of the textbook!

Academic integrity (1)

- All students are required to take the academic integrity quiz, this is a **requirement** to enter this course and pass this class
 - *Minimum* required score is the perfect score of 100
- In case of doubt –*any* doubt at all– I pass cases on to OSCCR for their evaluation

This is ***not*** a punitive measure:

sometimes “reeducation” is required, and NEU has an office that is specialized in evaluating, and hopefully dissolving these cases

Academic integrity (2)

- Assignment submissions are automatically and immediately scanned using a tool (plagiarism check) called “Turnitin”
 - The outcome of this scan is also immediatey available to students themselves (a plagiarism ‘score’)
 - Again, this is not intended for “policing”: but rather to help remind you when and where to provide proper attribution to the sources you are referencing
- *Quoting* sources is fine, but/and requires to elements:
 - Highlight the section that is quoted by using double quotes “<referenced text>”
 - Refer (“point”) to the origins of this quotation
- For academic references, APA standards apply

Late submissions

Late ($=< 7$ days) \Rightarrow max 50%

- Sometimes, special and unforeseeable events may happen in the student's life
 - If these occur, then exceptions may be requested

Teacher's leniency is proportionate to the ***timeliness*** of announcement

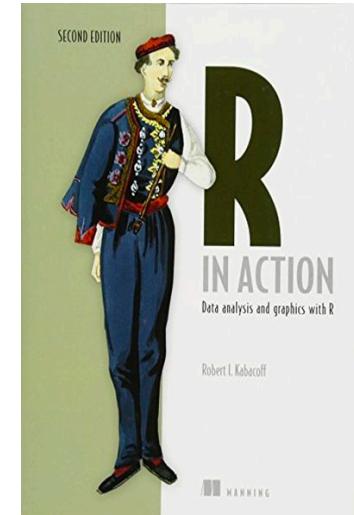
- If the events preclude you from completing your work on time, then the cause of your anticipated lateness needs to be communicated with teacher, who will evaluate on a case-by-case basis

You can do yourself (!) a favor by letting teacher know
as early as possible

Installing R & R-Studio

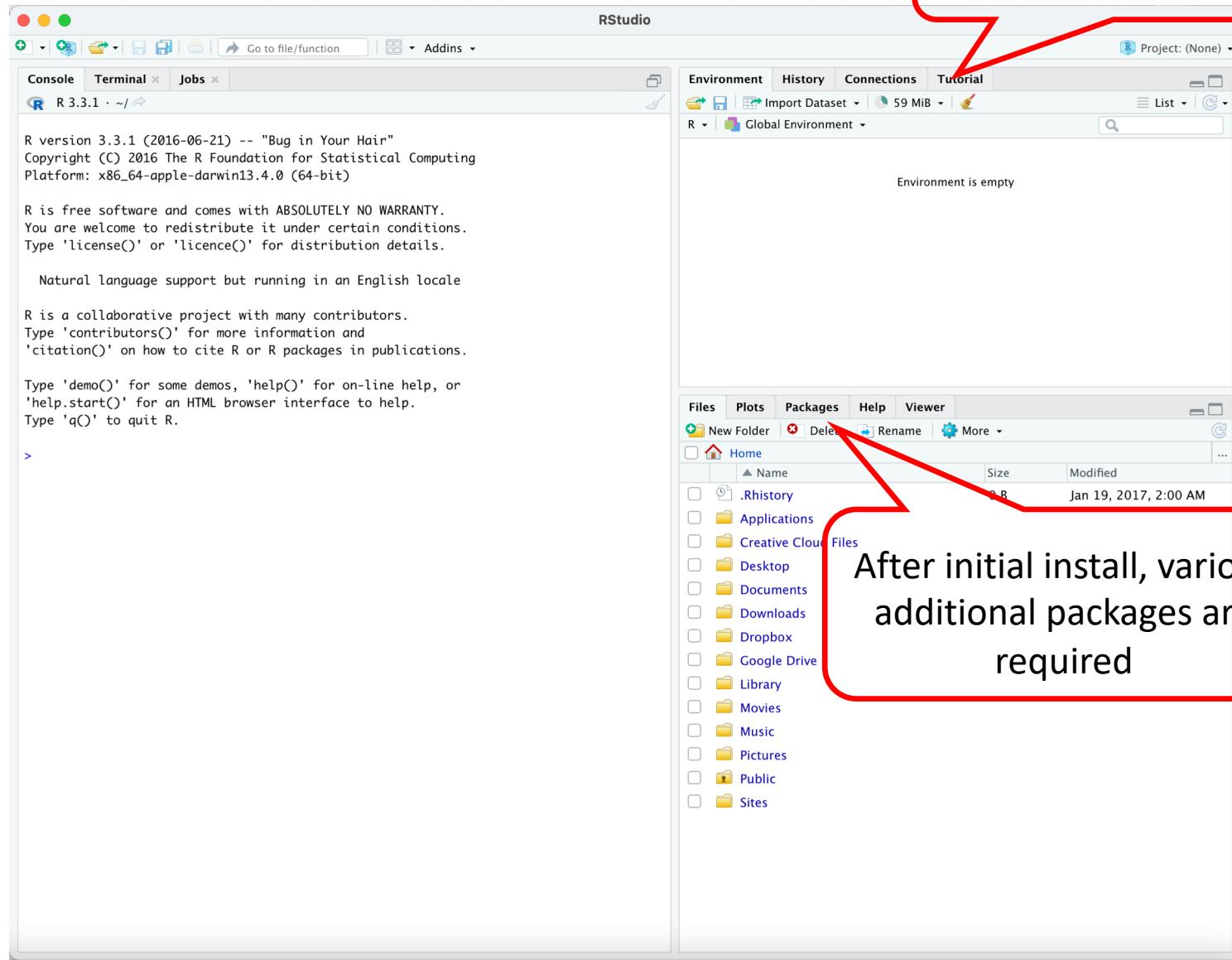
R-Studio & Kabacoff

- In week 1, you will find ungraded (!) assignments to “help” you get started with R & Bluman
- R-Studio software (<https://www.rstudio.com/>) is freely available for download:
<https://www.rstudio.com/products/rstudio/download/>
 - R-Studio IDE (“Integrated Development Environment”) is most commonly used for R



R-Studio

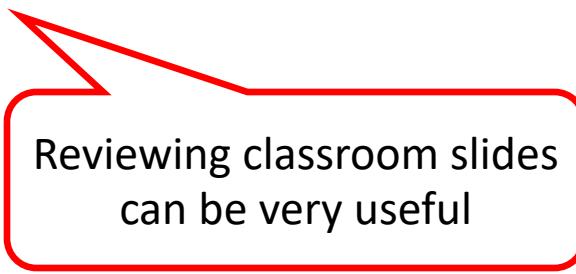
Student view may differ,
depending on OS



Teaching format

Session structure

- Every classroom session has the same format/structure:
 - Q&A + Administrative notes (“rules of engagement”)
 - Review materials, comment on submissions, and answer questions from last week
 - Elaborate on substantive materials, allow for Q&A, and formative (non-graded, interactive) assessment
 - Help students prepare for the upcoming week, by highlighting Discussion Forum requirements, and assignments: projects & R exercises
- Classroom slides are posted to Canvas immediately after class



Reviewing classroom slides
can be very useful

Discussion Forums

- Every week, there is a “Discussion Forum” requirement
- Discussion Forum contributions are worth 100 points, arguably the “easiest” 100 points you can earn in this course
 - Therefore: try *not* to drop any points here unnecessarily!
- Students can *only* see other submissions *after* posting their primary response
 - Primary response has a minimal (!) length of 250 words
- A minimum (!) of two responses are required, but more are allowed, each with a minimum of 80 words *substance*

Important:

Three posts need to be submitted on *three distinct* days!

R Practices

- Every week there are new and progressively challenging R practices: only in week 1 this is ungraded
 - Consider week 1 a lenient opportunity to “ramp up” – *not* an excuse to fall behind!!
- As course materials evolve and become more challenging, more packages and libraries will need to be installed
- Both statistical as well as data visualization tasks are an essential part of this course

Cease every R practice as an opportunity to hone
your visualization and data storytelling skills

Project assignments

- Every module has a “Project” requirement
- Each week, the project is graded with up to 100 pts
- “*Project management*” skills include:
 - Review
 - Quality control
 - Timely submissions
- Module 4-5 have “professional” themes, all others are reporting, visualization, or statistics focussed

Suggestions on how to succeed in this course

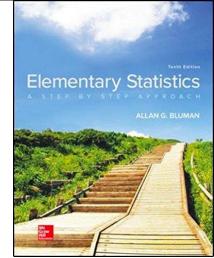
Lessons from the trenches
(AKA “best practices”)

Planning / time management

- The course ALY 6000 is fast paced, so it is imperative you do not allow yourself to “fall behind”
 - Experience shows that students that *do* fall behind, struggle to “catch up”, and tend to get far poorer grades
- The course is divided in many small(-ish) components to help you pace appropriately
 - In practice, you probably need to work on this course every single day to “keep up”, if only for a few hours

NB.: By faculty policy (!), there are ***no*** “do-overs” or opportunities for “make up work” in case of missed submission deadlines or disappointing grades

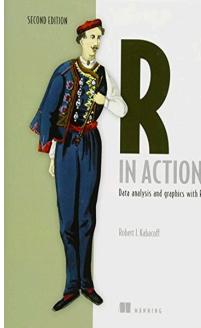
Recommended study method (1)



Bluman (10th Ed.)

- Step 1: review and internalize the chapter “Objectives”
 - For Chapter 6 these are on p. 311, **1 - 7**
 - For Chapter 7 these are on p. 369, **1 - 6**, etc.
- Step 2: read the chapter, until you find the answer/solution to objective **1**
- Step 3: **highlight the text** that provides the answer to **1**
- Step 4: go back to the chapter “Objectives”, find **2**
- Step 5: continue reading the chapter, until you find the answer/solution to objective **2**
- Step 6: **highlight the text** that provides the answer to **2**
- Rinse and repeat

Recommended study method (2)



Kabacoff / R Studio

- Practice, practice, practice
 - Programming skills are best acquired by running through exercises, *repeatedly*, and writing lots of code
 - Sometimes slightly tweaking your scripts can be (very) helpful to explore the program's behavior
 - Keep the table with “Help” functions on p. 11 close by!
- Beyond the basic steps of *operating* R, there are a ton of resources online you can tap into, like:
YouTube, (free) online training resources, R community,
<https://www.r-exercises.com/>, etc.

File submissions

- File names
 - In order to identify file contents, and the person submitting, it is helpful (**required!**) to make your file name *unique*, and *identifiable*
 - File names have a “structure”
`<contents>_<module>_<#>`
 - When you add a `_<name>` and `<date>` suffix (YYYYMMDD), this makes the file unique
- Test your “stuff” before you submit
 - A recurring source of disappointment is when students get poorly graded because their submissions were not “the latest” version, or an “incorrect” version
 - In order to avoid this, **test** (and **verify**) the files you intend to submit *prior* to sending them over!

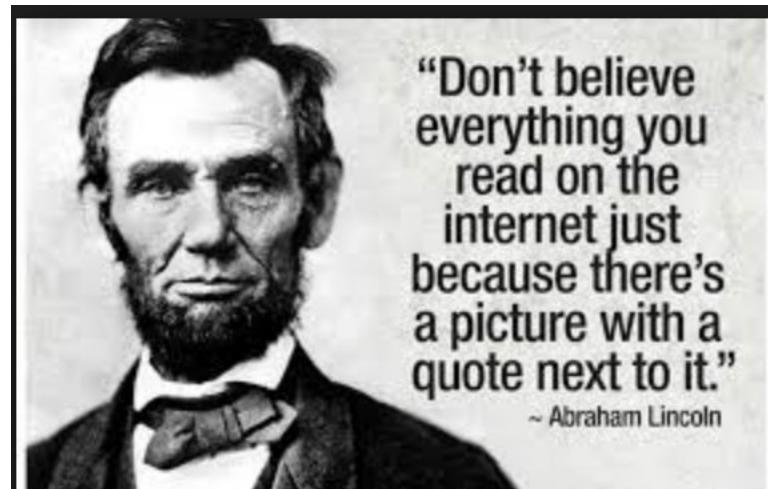
Timeliness

- Short version: submit on time
- Slightly longer version: plan well, so that you can verify your submissions before you submit, make any necessary corrections, and still submit on time
- If you *must* be late, then let teacher know; sometimes special circumstances may warrant an extension
 - N.B.: make sure you request an extension ***well before*** the deadline
- > 1" late = **late** ; 1" - 7 days late can still earn 50% of credit

Violin serenades *after* the deadline are best performed in private

Quoting internet sources

- Be aware that a “Google search” favors hits with high SEO, not necessarily most reliable and credible results
 - Companies that invest the most in SEO or SEM are not always putting in commensurate effort to improve their substantive content
- Always, *always* triangulate claims you read on the web



Preparation week #2

Requirements

- Academic Integrity quiz
- Reading Assignments
- Discussion board (100 pts) – due Sept 25th
- Module 1 R practice:
 - Installing R, then R-Studio
 - Kabacoff Chapters 1 & 2
- Module 1 statistical practice, section 1.1 to 1.4
- Module 1 Executive Summary Report 1 (100 pts) – due Sept 26th

Academic integrity quiz

- All students are required to take the academic integrity quiz, this is a **requirement** to enter this course and pass this class
 - *Minimum* required score is the perfect score of 100
- Students are allowed as many attempts as necessary to obtain the perfect score of 100
 - Results of less than 100 mean the student has *not* passed the test
- Academic integrity quiz does *not* count towards the final grade, it is an *entry requirement*

Reading assignments

- Bluman as well as the course shell contain introductory topics in Statistics
 - E.g.: graphical versus numerical descriptive Statistics, Inferential Statistics, Samples & Sampling methods, etc.
- Students are expected to complete these reading assignments **independently**
- The “Ask the Instructor” forum in Canvas is particularly well suited to ask teacher to explain concepts in the text

Make sure to “stay on top” of reading assignments, as an essential part of **prudent course planning**

Discussion board

- You *first* (!) need to post an original contribution (“primary post”) first, with a minimum (!) of 250 words
- A minimum (!) of two responses are required, each 80+ words, and posted on distinct, successive dates
- All contributions need to be substantive
 - For clarity: statements like “I agree”, “I like your post”, etc. **do not count as substantive replies.** Instead reason why you agree or disagree, and refer to outside sources (“references”) to justify your position
- Referring to other sources or posts, or previous classes, earns “brownie points” towards top grades (100 points) for so-called integrative learning

Discussion board Week 1

- Primary post by Thursday addressing the prompt. Responses for at least two of your classmates between Friday and Saturday (by 11:59 pm ET).
- Write 3-4 sentences about yourself that include key data points about yourself. Include both qualitative data and quantitative data. Please also respond to at least two of your classmates (by Friday if at all possible) with thoughtful follow-up questions or ideas on other data points to share
- Consider this an opportunity to make yourself familiar with quantitative as well as qualitative measurement. You might do well to create a graphical representation or infographic that highlights some noteworthy and memorable elements that help your peers relate to you, and remember you and your interests

Module 1 statistical practice

- Can be studied independently, concepts in e.g. Bluman are particularly well suited for questions in the “Ask the Instructor” forum
- Bluman problem solutions allow for self-assessment

Module 1 Executive Summary

- Dataset provided as part of the “Module 1 Project Instructions”
- Additional data in micro “student.csv” dataset
- This assignment involves *both* graphical as well as numerical descriptive statistics
 - Assignment also involves additional R “vcd” installations

Requirements for professional report development
will become progressively more challenging