

ANT 6937

DATA VISUALIZATION AND EXPLORATION

PLAN FOR TODAY

1. Introductions
2. Orientation
 - Syllabus and course website
 - R and RStudio
 - DataCamp
3. Complete your first in-class activity!

WHO AM I?

Dr. Fernando Campos

Department of Anthropology

Email: fernando.campos@utsa.edu

- Phone: 210-458-8635
- Office: MH 4.03.44

Office Hours:

- Tuesdays and Thursdays, 2:00 to 3:30
- By appointment
- Or take your chances and just drop in!



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.

Taught at Tulane University, New Orleans.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.

Taught at Tulane University, New Orleans.

Postdoc at Duke University, NC.



WHO AM I?

Dr. Fernando Campos

Born and raised in San Antonio.

Graduate school in Calgary, Canada.

Taught at Tulane University, New Orleans.

Postdoc at Duke University, NC.

This is my first year at UTSA!



WHAT DO I DO?

Dr. Fernando Campos

I'm a biological anthropologist, studying the behavioral ecology of wild primates.





Nicaragua

Costa Rica

Panama



Nicaragua

Área de Conservación
Guanacaste

Costa Rica

Panama





WHITE-FACED CAPUCHIN



MANTLED HOWLER



BLACK-HANDED SPIDER MONKEY



Kenya

Amboseli
National Park

Tanzania

AMBOSELI BABOON RESEARCH PROJECT



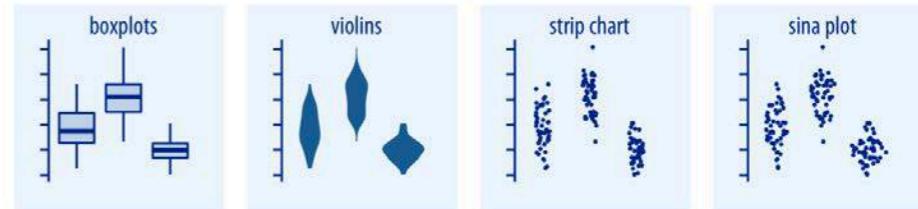
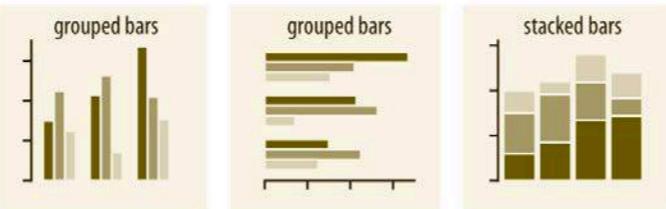
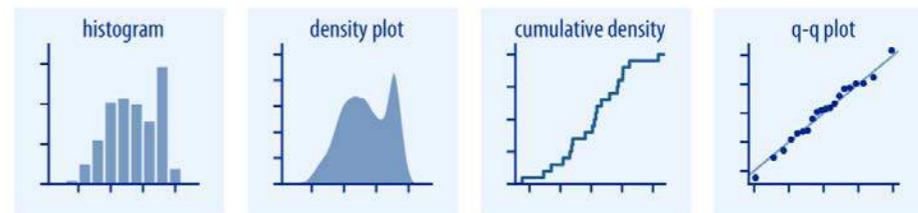
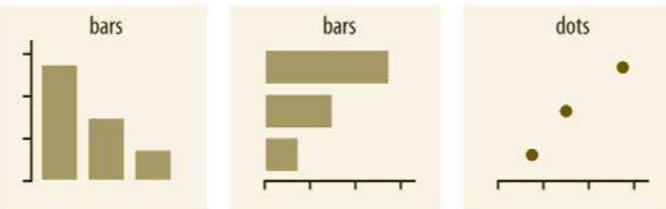




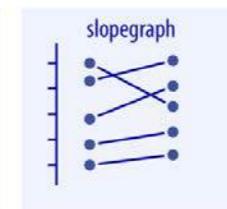
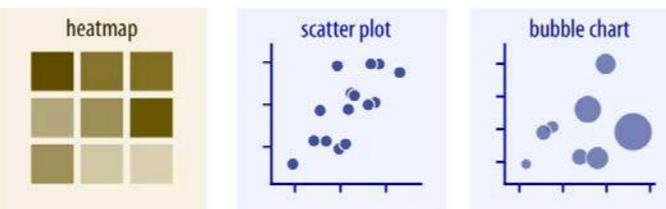
WHY AM I TEACHING
THIS COURSE?

DIRECTORY OF VISUALIZATIONS

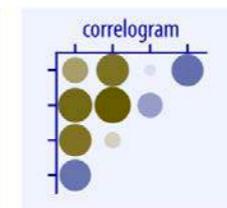
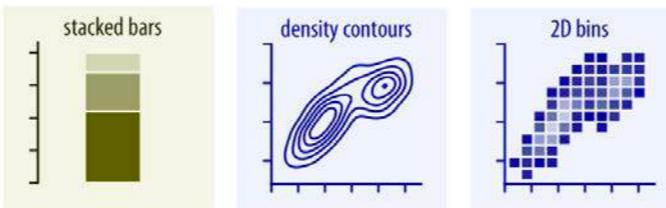
Amounts



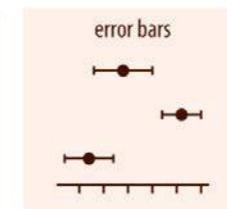
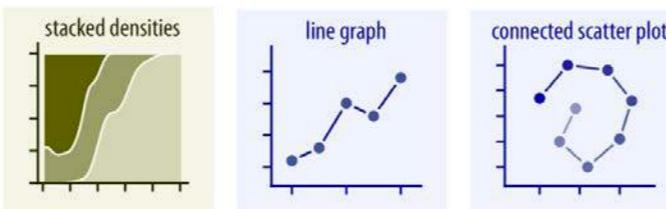
Proportions



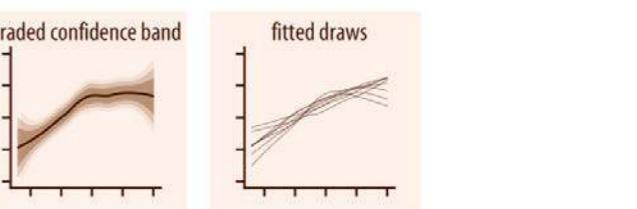
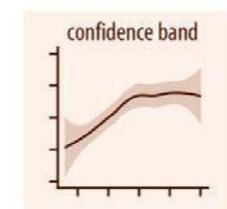
Distributions



Uncertainty



X-Y Relationships



YOUR TURN

1. Who are you?
2. What are your research interests?
3. What is your experience with data?
4. Do you have any experience using R, and if so, how much?
5. Is there anything specific that you hope to learn in this course?

**HELLO
my name is**

15:00

ORIENTATION

ORIENTATION

- Syllabus and course website
- R and RStudio
- DataCamp

COURSE WEBSITE

- <https://www.campos-lab.net/courses/ant6937-s2019/>
- Bookmark it!
- Go here for up-to-date information, schedule, assignments, readings, etc.
- I'll use Blackboard only for posting grades.
- Let's explore this a bit...

The screenshot shows a web browser window for the course website. The header reads "ANT 6937, Spring 2019". The top navigation bar includes links for HOME, DATA CAMP, CHEAT SHEETS, SCHEDULE, and POLICIES. The main content area features a background image of a hand-drawn style graph with various data series, including a green lightning bolt icon and a scatter plot. Below the image is a table with course details:

Instructor	Dr. Fernando Campos
Location	McKinney Humanities (MH) 3.02.52
Times	Thu 9:00 am - 11:45 am
Course	ANT 6937.002
CRN	34610
Office Hours	Tue/Thu, 2:00-3:30 pm, MH 4.03.44

Overview
This course is designed for graduate students in social or natural sciences who would like to gain experience in data visualization and analysis using R. It aims to be accessible to those who are completely new to programming, but also to provide new insights and skills to those who already have experience using R.

Approach
This course will use a mix of traditional and [flipped classroom](#) approaches, with new material introduced in reading assignments prior to class while class time will focus on applying those skills to data sets. The course will be activity-focused. Most concepts will be first introduced in assigned readings outside of class, leaving class time to focus on developing practical skills, completing activities, making progress on individual projects. We will cover topics in both data visualization as well as data manipulation. Students are encouraged to work collaboratively in and out of class. The [online course schedule](#) provides an overview of the topics covered as well as links to weekly readings, in-class activities, and any lecture material. This schedule is preliminary and always subject to change.

Texts
We will read sections from four openly licensed, online textbooks for this course. Hard copies of the books are *not* required, but some can be found at the links below.

- [Fundamentals of Data Visualization](#) by Claus Wilke. | [Hard copy](#).
- [Data Visualization: A practical introduction](#) by Kieran Healy. | [Hard copy](#).
- [ModernDive: An Introduction to Statistical and Data Sciences via R](#) by Chester Ismay and Albert Y. Kim. | [Hard copy](#).
- [R For Data Science](#) by Hadley Wickham and Garrett Grolemund. | [Hard copy](#).

Additional reading material will be linked from the [schedule](#).

At the bottom, there are links to GitHub, Textbook Wilke, Textbook Healy, Textbook ModernDive, Textbook R4DS, and a Creative Commons BY license logo.

ORIENTATION

- Syllabus and course website
- R and RStudio
- DataCamp

WHAT'S R VS. RSTUDIO?

- R is a programming language.
- RStudio is a nice user interface for R (an *integrated development environment* or IDE) that provides many convenient features and tools.



Engine



Dashboard

WHY R?



- R is the premier data analysis and visualization platform.

```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r

R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

WHY R?

- Pros:
 - Powerful, state-of-the-art
 - Used by many researchers—it is becoming a standard
 - Easy to extend, modify, and improve with add-on packages
 - Freely available for Unix, Windows & Mac
 - Programmable: if R can't do a particular task, you can program it.
 - Powerful plotting packages produce publication-quality graphics.



```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
> |
```

WHY R?

- Cons:
 - Command-based rather than interactive in the point-and-click sense (like SPSS & Excel)
 - Significant learning hurdle
 - Specific commands easy to forget
 - Documentation sometimes poor
 - Still evolving: backward-compatibility has been an issue
 - Slow at times when compared to other programming languages



```
● ● ● 1. R
Campos-Mac-Mini-2018:~ czx412$ r
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```



The screenshot shows the homepage of The Comprehensive R Archive Network (CRAN) at <https://cran.r-project.org>. The page features the R logo and navigation links for 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, and Contributed. The main content area includes sections for Download and Install R, Source Code for all Platforms, Questions About R, and What are R and CRAN?.

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 (2018-12-20, Eggshell Igloo) [R-3.5.2.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.

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. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.

Submitting to CRAN

To "submit" a package to CRAN, check that your submission meets the [CRAN Repository Policy](#) and then use the [web form](#).

If this fails, upload to <ftp://CRAN.R-project.org/incoming/> and send an email to CRAN-submissions@R-project.org following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

Note that we generally do not accept submissions of precompiled binaries due to security reasons. All binary distribution listed above are compiled by selected maintainers, who are in charge for all binaries of their platform, respectively.

For queries about this web site, please contact [the webmaster](#).

- Let's make sure we all have a current installation of R.
- <https://cran.r-project.org/>

- And a current installation of RStudio.
- <https://www.rstudio.com/>
- We'll get a crash course in how to use RStudio during the in-class activity.

Download RStudio - RStudio <https://www.rstudio.com/products/rstudio/download/#download>

R Studio Products Resources Pricing About Us Blogs

RStudio Desktop 1.1.463 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it here.

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

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.463 - Windows Vista/7/8/10	85.8 MB	2018-10-29	58b3d796d8cf96fb8580c62f46ab64d4
RStudio 1.1.463 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-10-29	a79032ba4d7daaa86a8da01948278d94
RStudio 1.1.463 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-10-29	8a6755fa9fae2bafce289df3358aaf63
RStudio 1.1.463 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-10-29	bc50d6bd34926c1cc3ae4a209d67d649
RStudio 1.1.463 - Ubuntu 16.04+/Debian 9+ (64-bit)	65 MB	2018-10-29	cf659db18619cc78d1592fefaa7c753
RStudio 1.1.463 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-10-29	742f0bad60dfeaa3281576e14ad6699e
RStudio 1.1.463 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-10-29	c7303067a0ca99deea7e427b856952d1

Zip/Tarballs

Zip/tar archives	Size	Date	MD5
RStudio 1.1.463 - Windows Vista/7/8/10	122.9 MB	2018-10-29	1eb1d7758bd4bf4bb68d4a7c3fe8d894
RStudio 1.1.463 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	90 MB	2018-10-29	ef9242b621d36c30de9d86b808840b41
RStudio 1.1.463 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	98.3 MB	2018-10-29	5944b3dd118cfba46f2a6c484d768324
RStudio 1.1.463 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.8 MB	2018-10-29	6bee446eeb4cc0e32967845e1f7ffe2b
RStudio 1.1.463 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	91.4 MB	2018-10-29	3f63725769b7b976fedb754eb2a19108

Source Code

A tarball containing source code for RStudio v1.1.463 can be downloaded from here

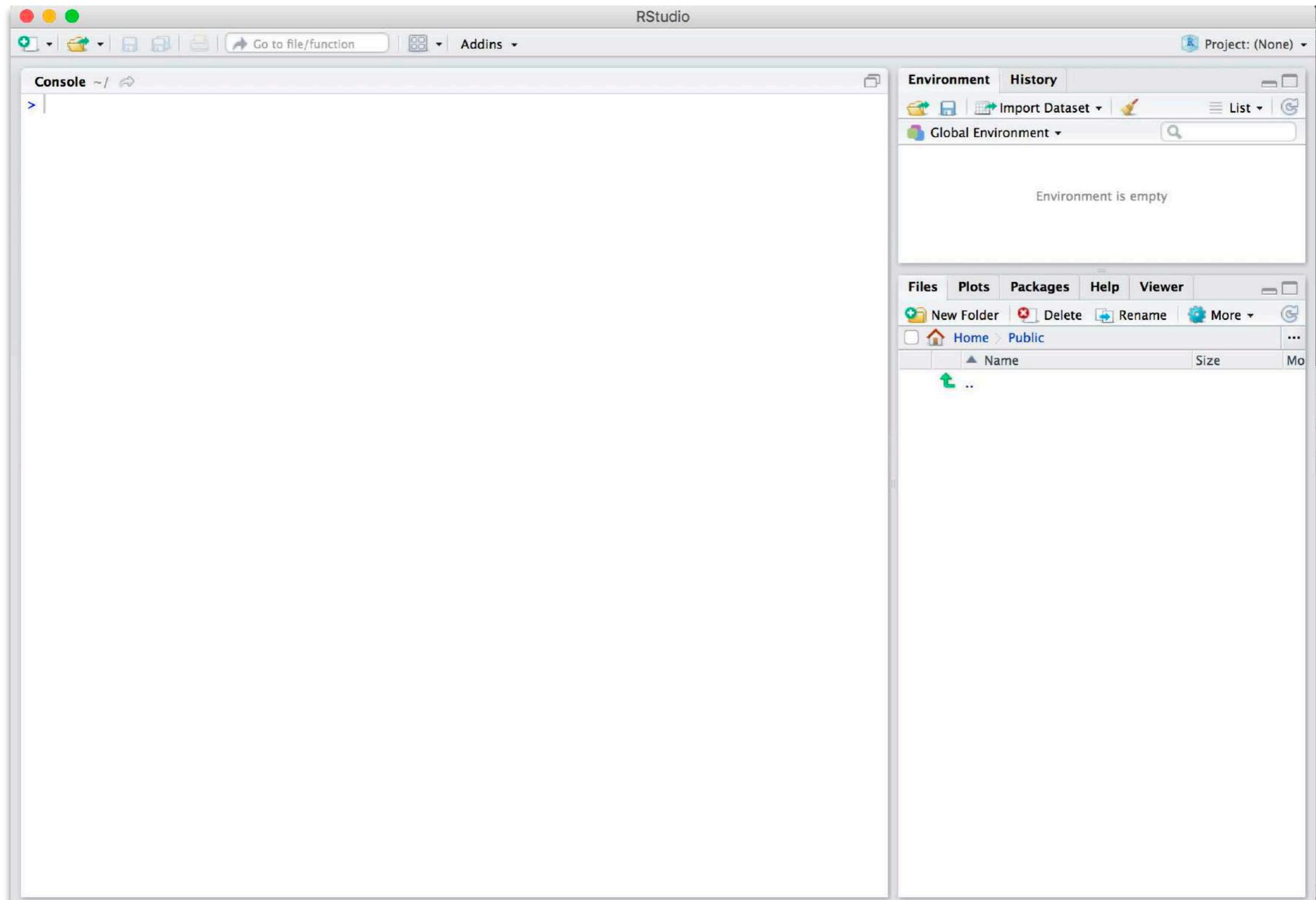
250 Northern Ave, Boston, MA 02210
844-448-1212
info@rstudio.com

DMCA
Trademark
Privacy Policy
ECCN

You can compile .R files just like .rmd files?!? 😜 The sneaky little notebook icon for doing this has been hiding... [twitter.com/i/web/status/1...](https://twitter.com/i/web/status/1212108840000000000) 12 hours ago

Try RStudio Server Pro for free!

Social media icons: Twitter, LinkedIn, Facebook, Email, GitHub



ORIENTATION

- Syllabus and course website
- R and RStudio
- DataCamp



DataCamp

- DataCamp is an online learning platform that focuses “data science,” which includes data visualization and data manipulation.
- It’s typically a paid service, but I have obtained licenses for you all.
- Let’s set it up.

The screenshot displays the DataCamp website with several key features highlighted:

- Top Navigation:** Learn R, Python & Data Science, Learn, Pricing, For Business, Sign in.
- Hero Section:** THE SMARTEST WAY TO Learn Data Science Online. Subtext: The skills people and businesses need to succeed are changing. No matter where you are in your career or what field you work in, you will need to understand the language of data. With DataCamp, you learn data science today and apply it tomorrow. Buttons: Start Learning For Free, python, R, SQL, spark, git, Shell, SPREADSHEETS.
- Create Your Free Account:** LinkedIn, Facebook, Google, Email address, Password, Create Free Account. Text: By continuing you accept the Terms of Use and Privacy Policy, that your data will be stored outside of the EU, and that you are 16 years or older.
- FOR INDIVIDUALS / FOR BUSINESS:** Tabs for different user types.
- Learn Anywhere, Anytime:** DataCamp's lessons are bite-sized so you can learn in a way that fits your schedule, on any device. Tracks conveniently order the courses so you can find what fits your needs at a glance. Testimonial: "I think it is obvious that DataCamp is the best platform for those seeking to learn data science." - CAMERON WHITE USED DATACAMP TO HELP HIM LAND A JOB AS A DATA ANALYST.
- Type conversion:** An exercise showing how to convert a float to a string using str(savings).
- Learn The Way You Learn Best:** You learn best when you get to apply what you learn quickly. That's why we've developed Courses that combine short expert videos with immediate hands-on-the-keyboard exercises. Why we have built Practice Mode with instant and personalised feedback on every exercise. And why we have developed DataCamp Projects based on real-world problems.
- Bottom Navigation:** python, R, SQL, spark, git, Shell, SPREADSHEETS.



DataCamp

- Go to “DataCamp” tab on the course website.
- In the future, you can use the link to the ANT 6937 DataCamp Dashboard.
- For now, click the invite link.

ANT 6937 DataCamp Dashboard

Throughout the course we will be using DataCamp, a platform that provides an interactive way to learn how to use R to create visualizations and carry out data analysis.

Invite link: https://www.datacamp.com/groups/shared_links/860db90029f1891286789d3ece4f95d11df40484

GITHUB TEXTBOOK: WILKE TEXTBOOK: HEALY TEXTBOOK: MODERNDIVE TEXTBOOK: R4DS

CC BY



DataCamp

The screenshot shows the DataCamp website's account creation page. At the top, there is a navigation bar with the DataCamp logo, a search bar, and links for "Sign in" and "Create Free Account". A yellow banner message reads "Please register or log in first before joining the group". Below this is a form titled "Create Your Account" with fields for First Name, Last Name, Email address, Password, and Phone. A "Get Started" button is at the bottom of the form. A small note below the form states: "By continuing you accept the Terms of Use and Privacy Policy, that your data will be stored outside of the EU, and that you are 16 years or older." To the right of the form, there is a "Sign in" link. The main content area below the form contains navigation links for LEARN (Courses, Skill Tracks, Career Tracks, Pricing, Become an Instructor), RESOURCES (Community, RDocumentation, Course Editor, Upcoming Courses, Support), PLANS (For Business, For Classrooms), and ABOUT (Stories, Careers, Contact, Press, Privacy Policy, Terms of Use). On the right side, there is a "DataCamp" logo and a brief description: "DataCamp offers interactive R, Python, Sheets, SQL and shell courses. All on topics in data science, statistics and machine learning. Learn from a team of expert teachers in the comfort of your browser with video lessons and fun coding challenges and projects." Below this is a section titled "ABOUT THE COMPANY" with social media icons for Facebook, Twitter, LinkedIn, and YouTube.

Create free account | DataCamp

Please register or log in first before joining the group

Sign in Create Free Account

Create Your Account

First Name

Last Name

Email address

Password

Phone

Get Started

By continuing you accept the Terms of Use and Privacy Policy, that your data will be stored outside of the EU, and that you are 16 years or older.

Already have an account? [Sign in](#)

LEARN

- Courses
- Skill Tracks
- Career Tracks
- Pricing
- Become an Instructor

RESOURCES

- Community
- RDocumentation
- Course Editor
- Upcoming Courses
- Support

PLANS

- For Business
- For Classrooms

ABOUT

- Stories
- Careers
- Contact
- Press
- Privacy Policy
- Terms of Use

[Download on the App Store](#) [GET IT ON Google Play](#)

© 2019 DataCamp Inc

DataCamp

DataCamp offers interactive R, Python, Sheets, SQL and shell courses. All on topics in data science, statistics and machine learning. Learn from a team of expert teachers in the comfort of your browser with video lessons and fun coding challenges and projects.

ABOUT THE COMPANY

f t in y

- Create an account using your @utsa.edu email address.

DATA CAMP

- Suggestion: turn off email notifications.

The screenshot shows the DataCamp website homepage. At the top right, a user profile for "Fernando Campos" is displayed, with a red box highlighting the profile picture and name. Below the profile, a sidebar titled "My Learning Progress" lists two recent course enrollments: "Intermediate SQL Server" (2019-01-16) and "Human Resources Analytics in R" (2019-01-14). A "Log Out" button is also visible in this sidebar. A green banner at the bottom encourages users to "Help us improve our courses by answering a few questions" with a "Get Started" button. The main content area features sections for "NEW COURSES" (with cards for "Intermediate SQL Server", "Hyperparameter Tuning in R", "Human Resources Analytics in R: Predicting Employee Churn", and "Interactive Data Visualization with plotly in R"), "DAILY PRACTICE" (with cards for "Joining Data in R with dplyr", "Introduction to Python for Data Science", "Introduction to R", and "Intermediate Python for Data Science"), and "GET REAL LIFE EXPERIENCE WITH PROJECTS" (with cards for "Introduction to DataCamp Projects", "Exploring 67 years of LEGO", "Dr. Semmelweis and the Discovery of Handwashing", and "Phyllotaxis: Draw flowers using mathematics"). A progress bar at the top right indicates "0 / 250 XP Gained", "0 Day streak", and "12 Hours Left". The URL in the browser address bar is https://www.datacamp.com/home.

DATA CAMP

- Suggestion: turn off email notifications.

The screenshot shows the 'Account Settings' page on the DataCamp website. A pink box highlights the 'Notifications' section in the sidebar. Another pink box highlights the 'Don't send me any emails' toggle switch, which is turned on (green). The 'Notifications' section includes options for Onboarding, Reminders, Course Launches, Newsletters, and Special Offers. The 'Don't send me any emails' section indicates that necessary emails like password resets will still be sent.

The screenshot shows the bottom section of the DataCamp website. It features a footer with several links: LEARN (Courses, Skill Tracks, Career Tracks, Pricing, Become an Instructor), RESOURCES (Community, RDocumentation, Course Editor, Upcoming Courses, Support), PLANS (For Business, For Classrooms), and ABOUT (Stories, Careers, Contact, Press, Privacy Policy, Terms of Use). To the right, there is a DataCamp logo with a brief description: 'DataCamp offers interactive R, Python, Sheets, SQL and shell courses. All on topics in data science, statistics and machine learning. Learn from a team of expert teachers in the comfort of your browser with video lessons and fun coding challenges and projects.' Below the description are links for 'ABOUT THE COMPANY' and social media icons for Facebook, Twitter, LinkedIn, and YouTube.



DataCamp

- Find links to DataCamp activities in course schedule under “Assignments”

The screenshot shows a course schedule for 'ANT 6937; Spring 2019' on the website https://www.campos-lab.net/courses/ant6937-s2019/schedule/. The schedule is organized by date and topic. A yellow warning bar at the top states: '⚠ PLEASE NOTE: Syllabus content will be updated frequently as we progress through the topics.' A red arrow points to the 'Assignments' column for the assignment listed on January 17, 2019.

Date	Topics	Readings	Assignments	Slides	Projects
R, RSTUDIO, AND THE TIDYVERSE					
2019-01-17	Introduction & Setup				
2019-01-24	R and RStudio Basics				
DATA VISUALIZATION					
2019-01-31	Visualization Basics				
2019-02-07	Visualizing Amounts and Parts of a Whole				
2019-02-14	Visualizing Distributions and Uncertainty				
2019-02-21	Visualizing Relationships and Change Over Time				
2019-02-28	Principles of Figure Design 1: Coordinate Systems, Color Scales, and Facets				
2019-03-07	Principles of Figure Design 2: Annotations, Themes, and Common Pitfalls				
2019-03-14	No class: Spring Break ☀				
DATA MANIPULATION					
2019-03-21	Data Import and Principles of Tidy Data				
2019-03-28	No class: AAPA Conference 🌎				
2019-04-04	Data Manipulation Basics				
2019-04-11	Grouping and Summarizing				
2019-04-18	Relational Data and Joins				
FINAL PROJECTS					
2019-04-25	Work on Final Projects				
2019-05-02	Work on Final Projects				
2019-05-05	Final Projects Due by End of Day (23:59:59)				

[GITHUB](#) [TEXTBOOK: WILKE](#) [TEXTBOOK: HEALY](#) [TEXTBOOK: MODERNDIVE](#) [TEXTBOOK: R4DS](#)

LET'S TAKE A SHORT BREAK

10:00