Prerequisite / Corequisite:

IST 687 or equivalent programming courses (Python, Java, SQL, C, C++, etc.)

Audience:

Students interested in data analytics and data science, with a focus on data/information visualization

Description:

A broad introduction to data visualization for information professionals. Students will develop a portfolio of resources, demonstrations, recipes, and examples of various data visualization techniques.

Additional Course Description:

Introduction to skills and techniques related to information visualization, through the R programming language and Adobe Illustrator. These skills include using data cleaning techniques, controlling the R graphics environment, developing custom plots, visually exploring data, using design concepts to visually communicate the story in the data, and discussing issues related to the ethics of data visualization. Conceptual themes will be presented alongside technical aspects of data visualization. Additional work and higher grading expected of graduate students.

Credits: 3

Learning Objectives:

After taking this course, the students will be able to

1. Use R to do basic data cleaning and preparation on a wide range of datasets
    a. Use functions to summarize and compare fields
    b. Find missing values
    c. Use subsets or filter data
    d. Retype data into correct format
2. Identify stories in datasets through exploration
    a. Use R to create appropriate rough plots to identify distributions and relationships in the data
    b. Use data subsetting and filtering to home in on questions of interest
3. Create rich visual artifacts that communicate data stories
    a. Identify the optimal type of visualization to minimize viewer cognitive overload and maximize image interpretability
    b. Enhance viewer cognition through context cues
    c. Use basic design principles to enhance viewer receptivity and convey meaning
    d. Use Adobe Illustrator to combine R data visualizations, design elements, and context cues into a single artifact

Bibliography/Texts/Supplies—Required:

Yau, N. (2011). *Visualize this: The Flowing Data guide to design, visualization, and statistics*. Wiley Publishing.

Yau, N. (2013). *Data points: Visualization that means something*. Wiley Publishing.

Bibliography/Texts/Supplies—Additional:

Additional readings will be provided as PDF files.

Requirements:

In order to meet the goals of the class (see above), we will use a combination of lectures to introduce topics and concepts, hands-on labs to introduce skills, group exercises and student presentations to enable peer-to-peer learning, and homework assignments to practice skills and gain deeper knowledge of course content. These are detailed below.

Grading:

Grade Scale:

A	96%–100%
A–	93%–95.9%
B+	90%–92.9%
B	87%–89.9%
B–	84%–86.9%
C+	81%–83.9%
C	78%–80.9%
C–	75%–77.9%

Assignments and Grading:

Assignment	Points
In-class labs	1 point per week
Quizzes	10 points each
Homework (Weeks 2, 3, and 4)	5 points each
Work-in-progress report	6 points
Viz-a-thon	3 points
Advanced topic presentation	8 points
Final project	25 points
Total	107 points

Your Final Grade Calculation

The assignment points you accumulate throughout the semester will be totaled and divided by 107. This calculated value will be matched against the above grade ranges to determine your letter grade.

In-Class Labs (1 point per week for the first 10 weeks):

Many of the lectures emulate a lab setting where the instructor writes code and explains it as he does along. The instructor will ask you to save some of the plots you make, which are **due at the beginning of the week's asynchronous lectures.** As it says above, labs are worth 1 point each. Note: no Lab assignments will be accepted or graded after our last "subject" class, specifically week 10.

Homework (5 points each) and Quizzes (10 points each):
These will be extensions of what we did in class or assignments out of the book. Homework may be in the form of quizzes, visualizations you create, or some of the other assignments listed below. Quizzes often lean heavily on the reading, and students who do not keep up with the readings often do not do well on the quizzes. Homework, quizzes, and progress report **are due midnight before the start of the next class**. If you turn in assignments late there are consequences. First, half of the assignment grade p o i n t s will be automatically deducted. Second, late assignments may not be graded until the end of the semester. As it says above, each homework is worth 5 points each and quizzes are worth 10. If a homework solution is covered in class and you have not turned in that homework you will not get any credit for that homework. Note: no HW assignments will be accepted or graded after our last "subject" class, specifically week 10.

Important: You may not receive credit if you do not follow the file-naming convention specified on the assignment sheet. You may not receive credit if your file is of the wrong type. Unless otherwise specified, you will always turn in plots as .pdf files and R scripts as .R files.

Attendance and Participation:
I do not directly grade on attendance or participation. Labs serve the function of taking attendance.

Advanced Topic Presentations (8 points):
Although this class is focused on creating visualizations using R, many other tools exist. Examples include D3, Tableau, and Gephi. In order to give students exposure to these and many other options, students will select, research and present on an "advanced topic." See the appendix for a rubric. Once you sign up for a topic you can start working on this assignment. **DUE: The slides should be uploaded before you present them.** The date will depend on which week the instructor assigns you to present. As it notes in the table, this assignment is worth 8 points.

Viz A Thon (3 points)
You will work in groups of 3 to go through the process of finding and visualizing a story in a data set provided prior to the start of class. You will turn in a mini-poster (8.5 x 11 inches, or 11 x 17 inches) as a pdf document. Speed and design are important! You must finish by the end of class. Do your best on design "wow" factor. Specific requirements are listed below

Final Project (25 points):
A large part of your grade comes from the final project, where you create and present a poster. The final poster project leverages skills developed throughout the semester, including the cleaning of data, the exploration of data with visualization techniques, data aggregation, simple design and information organization skills, and quality graphic presentation of data visualizations. Key deliverables leading up to the final project will be due throughout the semester to help the students stay on track for this major deliverable.

You must be present at the poster session at the end of the semester to get credit for your poster. The exact time and date of the poster session will be announced within the first few classes. Note that the requirements for posters change each semester. Past examples on 2U - Data Sets, Images, Readings 5 are provided for your reference but do consult the rubric. The work in progress report, which precedes the poster project, is worth 6 points and the poster is worth 25.

Course-Specific Policies (attendance, late work, makeup work, examinations if outside normal class time, etc.):

I do not directly grade on attendance or participation. Submitted lab work serves the function of taking attendance.

Late work will automatically lose half of grading points. So a 5-point assignment that is late may receive a maximum of only 2.5 points. Note: no assignments will be accepted or graded after our last "subject" class, specifically week 10 except the final poster project.

Additional Information:

Note on packages and external software: Almost everything we do is in R, but we do extend R by installing "packages." Packages can be installed within R. Below is a list of packages we will probably use during the semester. It would be helpful if you installed them before the lectures.

alluvial, animation, ape, aplpack, circular, devtools, extrafont, ggmap, ggplot2, hexbin, igraph, jsonlite, leaflet, lubridate, mapproj, maps, plotrix, png, pracma, raster, RColorBrewer, reshape, rgdal, rgl, riverplot, rnaturalearth, rworldmap, scales, scatterplot3d, shiny, stringr, TeachingDemos, Ternary, tidyr, treemap, vcd, vioplot, waffle, wordcloud, wordcloud2

Around Week 9, we do some simple animations. These require installing one of both of the following external software packages: ImageMagick and/or ffmpeg.

Academic Integrity Policy

Syracuse University's academic integrity policy reflects the high value that we, as a university community, place on honesty in academic work. The pilot policy in effect at the School of Information Studies defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. The pilot policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The pilot policy also prohibits students from submitting the same work in more than one class without receiving written authorization in advance from both instructors. Under the pilot policy, students found in violation are subject to grade sanctions determined by the course instructor and nongrade sanctions determined by the School or College where the course is offered. SU students are required to read an online summary of the university's academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during preterm check-in on MySlice. For more information and the pilot policy, see http://academicintegrity.syr.edu.

Disability-Related Accommodations

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), disabilityservices.syr.edu, located at 804 University Avenue, Room 309, or call 315-443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue "Accommodation Authorization Letters" to students as appropriate. Because accommodations may require early planning and generally are not

provided retroactively, please contact ODS as soon as possible. Our goal at the iSchool is to create learning environments that are useable, equitable, inclusive, and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, please meet with me to discuss additional strategies beyond official accommodations that may be helpful to your success.

Religious Observances Notification and Policy

SU's religious observances policy, found at supolicies.syr.edu/emp_ben/religious_observance.htm, recognizes the diversity of faiths

represented in the campus community and protects the rights of students, faculty, and staff to observe religious holy days according to their tradition. Under the policy, students should have an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance provided they notify their instructors no later than the end of the second week of classes through an online notification form in MySlice listed under **Student Services/Enrollment/My Religious Observances/Add a Notification**.

Student Academic Work Policy

Student work prepared for University courses in any media may be used for educational purposes, if the course syllabus makes clear that such use may occur. You grant permission to have your work used in this manner by registering for, and by continuing to be enrolled in, courses where such use of student work is announced in the course syllabus.

I intend to use academic work that you complete this semester for educational purposes in this course during this semester. Your registration and continued enrollment constitute your permission.

I intend to use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your written permission or render the work anonymous by removing all your personal identification.

Course Evaluations

There will be an end-of-course evaluation for you to complete this term, described below. This evaluation will be conducted online and is entirely anonymous. You will receive a notification from the Syracuse University Office of Institutional Research & Assessment (OIRA) department in your e-mail account with the evaluation website link and your passcode.

End-of-semester evaluation will be available for completion in Week 10 prior to your final poster presentation. This evaluation is slightly longer and is used to gauge the instructor performance and make adjustments to the course to ensure it meets our student needs.

We faculty work hard to do the best possible job when preparing and delivering courses for our students. Please understand that not only does the school use the course evaluations to make decisions about the curriculum to improve where necessary, they also use them to make decisions about faculty members. Please take the time and fill out this evaluation as your feedback and support of this assessment effort is very much appreciated.

School Library Media Program Assessment

The School Library Media Program is accredited by the Council for the Accreditation of Educator Preparation (CAEP) through the Syracuse University School of Education. As a part of that accreditation, the School Library Media Program must assess student performance on the competencies that correlate to program outcomes. The competencies that are assessed are identical to the items on your Competency Checklist. As a part of School Library Media Program planning, course-embedded assessments have been aligned with student competencies. For

CAEP reporting, each faculty member with competency-based, course-embedded assessments is asked to rate (1=Ineffective, 2=Developing, 3=Effective or 4=Highly Effective) candidates' performance on the respective competencies. This is the same rating scale students use when completing the competency checklist. Scoring is conducted for key assignments and not all assignments for a course.

What this means for you:

Your individual score is NOT a grade—it is part of an aggregate report. If a student is performing at an Ineffective or Developing level, a comment is submitted with the score, which is also aggregated. Individual scores and comments are not associated with specific student names.

Weekly Topics and Assignments

Week	Topic	Required Reading and Assignment
1	What Is Data Visualization? What Is R? Learning outcomes—students will be able to • Differentiate between information visualization and other prominent forms of visualization • Describe the two main purposes of information visualization: exploration and communication • Describe the seven basic steps of visualization • Describe the important features of the RStudio IDE • Create variables in R • Create simple single-variable plots in R, such as pie and bar charts and histograms • Iteratively build up increasingly complex plots • Use R's Help system to look up the available parameters for plotting functions • Construct file name strings • Open data files	**Week 1 Lab—Asynchronous**: R and basic plots. Follow along with the video and turn in the plots and script. **Week 1 Lab—Live Code-Along**: RStudio environment and loading data. **Sign up for advanced topic presentations.** **Readings**: • Visualize This: Chapter 1 • Visualize This: Chapter 2 (Read only from page 22 through 27. Skip the rest of the chapter.) • Ben Fry, *Visualizing data: Accessibility score*, Chapter 1, The Seven Stages of Data Visualization. • Wiley and Pace, *Beginning R: An introduction to statistical programming*, Chapters 1 and 3. **Quiz Week 1**: **(10 points)**. Covers readings, lectures, and labs.

2	**Data and R** Learning outcomes—students will be able to • List sources of freely available data • Describe the process of data exploration • Identify context markers in visualization • Differentiate between common data types • Use R functions to explore and clean data • Use R to retype, subset, and filter data • Use R to aggregate and group data • Create rough data exploration plots • Describe ways to make comparisons with visualizations • Describe ways to identify and show relationships in data • Differentiate between single- and multidimension plots	**Week 2 Lab—Asynchronous:** Exploring data in R. **Sign up for advanced topic presentations.** **Readings:** • Wiley and Pace, *Beginning R: An introduction to statistical programming*, Chapters 2 and 4 (PDFs). • Data Points: Chapter 1. • Data Points: See Chapter 4, section on Distributions, pp. 189–199. • Visualize This: Chapter 4.

Week	Topic	Required Reading and Assignment
	• Make simple multidimensional plots in R • Identify the appropriate plot type for a given set of data • Identify questions that might be answered with the data	• Visualize This: Chapter 3 is optional. **Week 2 Homework**: **(5 points)** See details below. **Quiz Week 2: (10 points)**. Covers readings, lectures, and labs.
3	**Graphic Design Principles: Color** **Learning outcomes—students will be able to** • Describe the use of contrasting and harmonious color in visualization • Describe how hue, saturation, and value combine to make a color • Use online tools to choose and create color schemes • Use R's color setting and transformation functions • Use color to provide visual cues in visualizations • Identify elements of a visual artifact that make it compelling • Interpret the meaning(s) of a data visualization • Use Illustrator to make simple modifications to R plots • Add context elements to a data visualization • Explain the difference between raster and vector graphics	**Week 3 Lab:** Working with color in R, Adobe Illustrator introduction, adding images to plots. **Readings**: Data Points: Chapters 2 and 3. Data Points: Chapter 4 (For now, skip pages 165–176; we'll cover that later.) Few: Chapter 3. **Week 3 Homework: (5 points)** See details below.

Week	Topic	Required Reading and Assignment
4	**Graphic Design Principles: Typeface and Layout, R Plot Area Control** **Learning outcomes students will be able to** • Describe how typeface and layout work together to create a visual hierarchy • Describe how visual hierarchies direct viewers' attention • Explain how lines, gutters, grids, and colors can be used to highlight visual elements • Critically assess example posters, and discuss effectiveness of design elements • Set and use different layouts in R • Set and use different kinds of fonts in R • Use Adobe Illustrator to carefully lay out text, plots, and other graphic elements of a layout • Export plots to Microsoft Word and other Microsoft products	**Week 4 Lab:** Layouts and fonts in R and details in Illustrator. **Readings:** Few: Chapter 4, Analytic Interaction and Navigation. Data Points: Chapter 5, Visualizing With Clarity. Visualize This: Chapter 4 (Illustrator parts). Visualize This: Chapters 6, Visualizing Relationships. **Week 4 Homework:** VT, Chapter 4, Illustrator, and Chapter 6. See detailed description below. **Quiz Week 4: (10 points)** Covers readings, lectures, and labs. **Note:**
5	**Working With Social Media Data: Twitter** **Learning outcomes—students will be able to** • Format, work with, and plot data with dates • Clean malformatted string data • Clean and parse hashtags • Use different data transformations to scale skewed data • Make word clouds • Develop complex categorical data plot, like alluvial, treemap, and river plots in R • Use example plot data as an approach to understand how to make new plot types	**Week 5 Lab:** Working with Twitter data. **Readings:** Visualize This: Chapter 7, Spotting Differences. **Assignment:** Start working on Work-in-Progress report (due Week 7). See details below.
6	**A Grammar of Graphics: ggplot2** **Learning outcomes—students will be able to** • Identify the building blocks of the grammar of graphics • Distinguish between aesthetics, geometry, and the other building blocks and what they do • Know the difference between setting and mapping aesthetics • Build basic and some complex plots with ggplot • Make informed decisions about when to use ggplot and when base plots or other packages are more appropriate	**Week 6 Lab:** ggplot. **Readings:** Mathison, The credibility of image-based research and evaluation. Sturken, Practices of looking. **Optional Reading** *R graphics cookbook* (PDF). **Assignment:** Finish Work-in-Progress report for next week. See details below.

Week	Topic	Required Reading and Assignment
	• Describe R's memory model and the implications for using ggplot with large datasets	
7	**A Grammar of Graphics: Maps, Work-in-Progress, and Ethics Discussion** **Learning outcomes—students will be able to** • Use different R packages to make maps • Geocode address data • Plot points on a map • Create choropleth maps at country, state, and county level • Simple regular expressions for cleaning data • Using shape files and geoJSON for map plotting • Discuss the difference between raster map tiles and vector shapes • Give constructive feedback on layouts, colors, fonts, and other design elements • Identify ethical concerns around visualizations	**Week 7 Lab:** Maps! **Readings:** Data Points: Chapter 4. pp. 165–176 (spatial data). Visualize This: Chapter 8. **Optional Reading** *R graphics cookbook* (PDF). **Assignment:** Work-in-Progress report due **(6 points)**. Prepare for your presentations. See details below. **In-Class Exercise:** Participate in group feedback session.
8	**Visualizing Social Networks and Advanced Topic Presentations** **Learning outcomes—students will be able to** • Describe key elements of a social network • Describe the concept of centrality in a network • Identify a few structural measures • Differentiate between matrix and linked list social network data • Visualize nodal attributes • Visualize network structures • Describe the meaning of a network layout • Describe advanced visualization tools used in the market place • Have knowledge of a range of additional R packages used in data cleaning and visualization not otherwise covered in class • Apply design skills learned in class to slide deck creation	**Week 8 Lab:** Social network visualization. **Assignment:** Prepare for your presentations Work on final poster (see rubric and instructions below). **Quiz Week 8: (10 points)** **In-Class Exercise:** Advanced topic presentations (8 points).

Week	Topic	Required Reading and Assignment
9	**RGL (3-D Visualization), Animation, and Advanced Topic Presentations** **Learning outcomes—students will be able to** • Create simple points and lines in 3-D space • Create a 3-D social network model • Control aspects of the 3-D environment, like lighting and simple materials • Create a simple data animation in 3-D space • Create a simple scene in 3-D space • Manage R packages • Describe advanced visualization tools used in the marketplace • Have knowledge of a range of additional R packages used in data cleaning and visualization not otherwise covered in class • Apply design skills learned in class to slide deck creation	**Week 9 Lab:** Visualizing in 3-D and making animations. **Assignment:** Prepare for your presentations. Work on final poster. See details below. **In-Class Exercise:** Advanced topic presentations (8 points).
10	**Shiny: Making a Simple Interactive Dashboard in R** **Learning outcomes—students will be able to** • Identify the fundamental functions for a Shiny app • Discuss the general architecture of a Shiny app • Build a simple Shiny application useful as a portfolio piece. • Demonstrate knowledge of the process of creating a visual artifact • Work in a group to create a mini-poster from an unknown dataset	**Week 10 Lab:** Shiny **Assignment:** Work on final poster. See details below. **In-Class Exercise:** Viz-a-thon (3 points).
11	**Poster Day** **Students present their posters.**	**Poster Presentations**

Weekly Lab's
During the weekly asynchronous sessions, you will be asked in the recorded lecture to re-create specific code as well as execute that code. Your weekly lab deliverables are a pdf file that shows the plots rendered via the code you were asked to create and the R code/script.

Week 2 Homework: (5 points)
This homework assignment has three parts. In the first part, you will create charts from exercises in *Visualize This*. In second and third parts, you will create plots to examine simple distributions and multidimension plots, most of which will require some form of grouping. The second and third parts are very similar to the labwork but with a different dataset.

Part 1
Use the *Visualize This* book and make the plots below. Only do the R part. (*Do not complete Illustrator exercises yet.*) A key part of this assignment is to make the plots as close to the same aspect (height-to-width ratio) as the plots in the book. So, if the plot is almost square in the book, yours should also be almost square. Size the plot window in R before exporting your plot. The plot width and height should be proportionally the same as in the book. This will be part of the grade. Match the examples in the book as completely as possible. (The dates won't quite match in one of the plots, in which case, include all the dates.)

- Bar chart: Figure 4-11
- Stacked bar chart: Figure 4-22
- Scatterplot: Figure 4-28
- Time series: Figure 4-34
- Step chart: Figure 4-43

Tips: Resize the chart output window BEFORE you save it as a PDF file to achieve proper proportions. Note that the plot width and height should be proportionally the same as in the book. This will be part of the grade. Create a script in a document window, and then use the console window to open/run it. This will enable you to reuse code and troubleshoot if there are problems. You will lose points if the plot is not a PDF file. Make sure you can open the file in Adobe Acrobat before turning it in.

Part 2: Simple Distributions
Download art.csv (from the Data and Images section) to your (or a lab) machine. Using R, make four plots to answer the following questions. Here, we are interested in looking at different ways to look at the same data. Use the par() function to put all four plots in the same plot space (like we did in the lab). Also, give the plots titles and *x*- and *y*-axis labels, use colors that you like, and, if you are using a plot with points, use a symbol you like (pch). In other words, customize these plots to show me that you know how to modify different elements of the plots.

- What is the distribution of total.sales for the whole dataset? Provide two different plots that show two different ways of showing distribution. Title your plot(s): Distribution of total.sales
- Next we want to compare the distributions of subsets of total.sales. Use a third type of distribution plot (different from what you used for the question above) for both of these plots.
    - What is the distribution of the totals sales for drawing paper? Title your plot "distribution of the totals sales for drawing paper"
    - What is the distribution of the totals sales for watercolor paper? Title your plot: "distribution of the totals sales for watercolor paper"

Part 3: Grouping and Multidimension Plots
Using the art.csv dataset again, answer the questions below. This work will require that you make plots that show relationships or allow comparisons. You will need to use some of the grouping functions and data sub-setting we used in the lab. Don't forget to set stringsAsFactors = FALSE, or it will affect the results of your plots. Also, put all three plots in the same plot space again, and customize colors and other parameters in some interesting way.

- Is there a relationship between the unit price of art goods and their units sold? If so, what kind of relationship is it? Indicate which plot answers this question.
- Does the art company sell more units of drawing paper or watercolor paper? Indicate which plot answers this question.
- Does the art company bring in more money (income) selling drawing paper or watercolor paper? Indicate which plot answers this question.

Submit

A single PDF file with all the plots. For Part 1, each plot should be its own page. Then there should be two more pages, one for Part 2 and one for Part 3. <LastName_wk2hw.pdf>. Also, turn in your script that you created to make these plots with the file name: <LastName_wk2hw.R>.

Week 3 Homework: (5 points)

Visually Describe a Dataset

The purpose of this exercise is to find a dataset that is interesting to you and then use what you have learned so far to visually describe a dataset with R plots. This means showing a few (3) single-dimension plots as well as at least one multidimension plot. One of the hardest parts of this assignment is going to be finding the right data. You will turn in a PDF file with the plots and your R script. Make sure your plots are vector graphics (saved as a PDF file). Each plot should be on its own page. Note that work done in this assignment can be used in both the Work-in-Progress report (coming up) and the Final Poster.

Data
You need to find a substantial dataset for this project. Your dataset must score at least 100 in the formula below:

(NumberOfColumns * 4) * (NumberOfRows/100) >= 100

Examples:

A dataset with 5 columns and 10,000 rows would score: (5 * 4) * (10,000/100) = 2000
A dataset with 17 columns and 3,000 rows would score: (17 * 4) * (350/100) = 238
A dataset with 5 columns and 500 rows would score: (5 * 4) * (500/100) = 100

The point of this is to make sure you have a large enough dataset to work with. You need enough columns to look at how different variables might be related. You need enough rows to show that you can work with a decent-sized dataset.

Provide the following information about your data:

- Description – what does the data set represent, 1 or 2 sentences
- Provide the output of the str function on your data set
- Provide the calculation notes above using your data set, specifically: (NumberOfColumns * 4) * (NumberOfRows/100)

Descriptive Plots
Your data may have any types. For example, you might have a mix of continuous, text, and categorical data columns, or you might have several columns of just continuous data. The type of the data should dictate what kinds of plots you make.

To visually describe the data, first, make some plots in R that show how the data are distributed. For continuous data, this might be histograms, box plots, or density plots. For categorical data, a good choice is a bar chart that shows the frequencies of different categories. A pie charts works for this, too. Include at least three of these single-dimension plots. If you tell me your data have only three columns, I will probably say to find a different dataset. Next, make a multidimension plot using at least two columns, though as you have seen in labs, some two-dimension plots can take three different columns to make. Indicate which plots are distribution oriented ie single dimension and which plot is multidimensional.

Note that in this assignment the plots and their context text should be all I need to get a sense of what your data are about. The plots don't need to be "pretty" unless you want them to be, but they do need context items that help me understand your data. All plots should have a title and possibly context text under the title so I know what the data are about. Include titles and axis labels in the plots. All plots should list the source in the lower-right corner in small (but readable) text.

Format of the PDF File
The file should be a single PDF file. The first page should include a title, your name, the source of the data, and a table of contents of the plots. Include your plots on the following pages. You can use functions like par(mfrow = …) to have more than one plot per page.

Upload this PDF file with the file name: DataReport-<your name>.pdf

Also upload your R script.

Week 4 Homework (5 points)

Part 1: Modifying Plots with Illustrator
Go back over VT, Chapter 4. Complete the Illustrator portions of the six chart-building exercises in this chapter. Note, you already made these plots in a previous homework, so you might be able to just modify those. Match the examples in the book as close as you can. Your grade will depend on how close your version matches the book, including the plot aspect (width-to-height ratio).

Bar chart: Figure 4-5
Stacked bar chart: Figure 4-21
Scatterplot: Figure 4-25
Time series: Figure 4-40
Step chart: Figure 4-42
LOESS curve: Figure 4-47

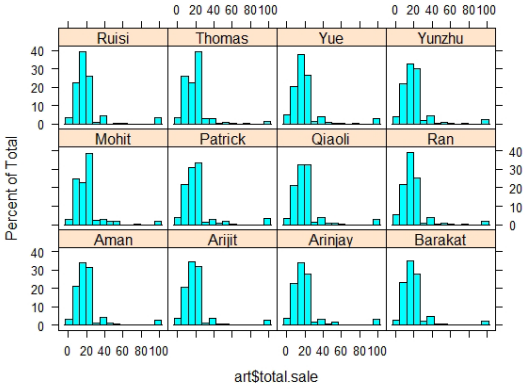
Because the grade will be based on how closely elements of the plot match how they look in the book, pay attention to things like the size and position of text, the proportions of the plot, and the colors of the background, lines, and text.

Begin by cleaning up the layers that are automatically generated by R when it creates the PDF file. Discard anything that you do not need. If you are having trouble selecting an object, try to right-click and choose **Releasing Clipping Mask** or **Ungroup**. This should release the object and allow you to edit it.

Save the plots as a PDF file from Illustrator.

Part 2: VT, Chapter 6, Plots
Submit exercises in VT, Chapter 6. (Note: The data for the first four plots are the author's original .csv. Chapter 6 has detailed instructions on downloading those.)
Scatterplot matrix, Figure 6-9
Bubble chart, Figure 6-15
Histogram, Figure 6-24
Density plot, Figure 6-32

Then, make one or two small multiple dimension plots using data of your choice—see Figures 6-38 and 6-40. Note, you either need the lattice or ggplot2 packages to make small multiple plots. You can use the art or sales datasets if you like. If you are not using art or sales dataset, please identify/describe your data set and where you sourced it from. Here is an example using the lattice package with the art dataset.



Please note that Adobe Illustrator is not required for completing Part 2: VT Chapter 6, including the last two questions. Formatting all your plots in R is sufficient. However you are welcome to use Illustratot if you would like.

Combine all plots into a single, multipage PDF file. Turn in the PDF file and your R script with the naming convention NetID_week4.pdf/NetID_week4.R.

Work-in-Progress Report—Due Week 7 (6 points)

Submit a PDF file of your Work-in-Progress report that includes the following:

Section 1 (1 point):

- Your NAME, SECTION, and TITLE.
- Dataset description of two sentences.
- Your compelling story of two sentences.
- Your audience (who would care about it) of two sentences.
- Your Questions 2 to 4. Include the page number of the plot (below) that answers the question.

Section 2 (another 1 to 3 pages) (1 points)

- A full-page sketch of how you might lay out your poster. You could have more than one sketch if you want. Just add one page for each sketch. You should refer to the week material where Graphic Design Principles were discussed specifically Typeface and Layout. Your sketch needs to reference the plots mentioned in Section 3

Section 3 (as many pages as you need) (1 points)

- At least three plots, but I suggest more like five. No upper limit. You should have enough plots to answer your questions from page 1, but you can have extra plots. The plots can all be on a single page or on multiple pages and should be used as examples when you get feedback.

Note: Sections 1,2,3 are due the evening before the WIP Report is discussed and critiqued by your team mates

IN-CLASS PARTICIPATION WORTH 3 ADDITIONAL POINTS.
Come to class prepared to share your report with a group.
Write down who was in your group, and take notes on the feedback you got.
In-class participation notes are due before the end of this scheduled class,, notes received the following or subsequent days will loose points

As part of what you turn in, include a photo or a word doc of your notes with the names of your group members.

For the poster sketch, you can make a mock-up in Illustrator, or you can you draw a sketch on paper and scan or photograph it and include it in the package.

Advanced Topic Presentation (Weeks 8 and 9) (8 points)

Presenting information on slides is a kind of information visualization. These are **8 to 12-minute** presentations on an advanced visualization topic. Each student should sign up for one of the topics (see below) to present. Note: presentation time noted previously may change due to the number of students enrolled in a class. You will be notified as to the exact amount of time you have.

Check with the section instructor about how they want you to sign up. But sign up early! No duplicate topics accepted. Your topic needs to be selected from the published topic list posted as a Google doc in the files area of 2U

Rubric

1 point: Your name is clear and bold on the title and closing slide.
1 points: Your presentation is within the allotted time.
4 points: You must teach us something about selected visualization tool via examples you created vs vendor created sample. You might want to consider using one of the data sets used in async or sync lectures ie sales, art, to demonstrate package capability
- Features - summary
- Functions / Capability– use a data set to demonstrate
- Your assessment of learning curve

1 point: Presentation is practiced—not too fast, smooth delivery, no technical glitches.
1 point: Slides are well designed without too much text.

***Note*:**
- You can earn from 1 to 5 NEGATIVE POINTS on your own presentation for talking during others' presentations.
- You will lose 5 points automatically if you are not ready on the day you are assigned.
- Turn in a PDF file of your slides, but you can present using a Power Point slide deck.

IMPORTANT:
- Demos are risky. They often don't work as planned, and the time counts against you. They often take more time than expected. They are one more moving part that can go wrong in so many ways! Demo's are risky however let's discuss.
- If you are showing us a web page, try it out on multiple machines, particularly on a lab machine. Generally avoid.
- No videos. No videos. No videos. Unless you made it. This is about your work, not someone else's.
- A single slide takes about 1 minute to present. After the title slide, plan to have seven or eight slides of content and one "Thank You" slide (which takes 0 minutes).
- Practice your presentation at least three times, out loud. Better if you do in front of others or a mirror.
- Avoid too much text on a slide. Better to use a few small images that remind you what to talk about.

List of topics (others may be approved by the instructor)

Topic

Carto
Chart Studio (plotly)
Chart.js
ChartBlocks
D3
Dash
DataHero

Dataiku

DataWrapper
Gephi
Google Charts
Infogram
InstantAtlas
Klipfolio
Looker
Panopticon
Plotly
Power BI
Qlik
RAWGraphs
Redash
SPLUNK
Superset
Tableau
Tangle
VISME
Visualize
Free
iGraph

Viz A Thon

Viz-a-thon: You will work in groups of 3 to go through the process of finding and visualizing a story in the data (link below). You will turn in a mini-poster (8.5 x 11 inches, or 11 x 17 inches) as a pdf document. Speed and design are important! You must finish by the end of class. Do your best on design "wow" factor.

1. Explore the data using R functions we have learned in class: dim(), colnames(), str(), fix(), head(), tail(), table(), tapply(), aggregate and so on
2. Identify 2 related questions that the data can answer that focuses on how some data are related to other data.
3. Use functions like table and aggregate to get the data ready for plot as needed.
4. Create two visualizations, each with 2 or more dimensions, that answer your questions.
5. Create two or more small single dimension plots that show the distribution of the data you are using in some way.
6. Create the mini-poster (8.5 x 11 inches, or 11 x 17 inches) document in Adobe Illustrator, but save and turn it in as a pdf!
7. Import each visualization into your Illustrator document and add
   - title
   - all team member's names
   - context text such as what the data is and what your two questions are
   - data source
   - your 4 or more plots with context, color, labels and so on.

8. Make sure you spend time with the design. Your mini-poster should be a compelling data story with consistent colors and good layout.

Note that on this mini-poster your small plots might be very small (example 1.5 by 1.5, or 2 by 2 inches), while for the larger ones you could make one as large as 4 by 4 inches.

Each larger visualizations should:

- Relate 2 or more dimensions of the data
- Be related to each other such that the two tell a story from the data

Data (TBD)

- Data file: Instructor will provide

Submit - Each member of the group needs to upload:

- The R script that opens, interrogates and plots the data
- The finalized visualizations arranged on a single page pdf (8.5 x 11 or 11 x 17).
- **Have one team member post the team's solution to the wall**

Rubric

- All elements listed above are present (2 multiDim and 2 singleDim plots, required text, correct size of pdf): 1 pt
- You have 2 questions from the data and your plots answer the questions: 1 pt
- Wow factor. Use colors, layout and fonts to enhance your story: 1 pt.

Final Poster Project: 25 points

The Instructor Should indicate the time of the final poster session.

For your final project, you will be creating an information visualization poster and presenting your work during the poster session that is open to the iSchool community and employers.

See example posters

I recommend watching the short video below, which provides some tips on how to present a poster.

https://www.youtube.com/watch?v=0ozwCEeaVWE

YOU MUST UPLOAD YOUR POSTER AS A PDF TO GET CREDIT. It will be considered late if not uploaded before poster session.

The procedure you will follow will be very similar to the visualization report you created earlier in the semester.

Go over the rubric carefully. You will be graded against the rubric (below). Also look at the examples, but note that the requirements of the project change, so the examples may not match the rubric.

The question I get asked most: Do you need to use R for the plots in your final poster?

Answer: No! Only if you want full credit should you use R for all the final plots. Note, you can add extra plots if they help you tell the data story, but the rubric plots must be made with R.

DATA

You need to find a substantial dataset for this project. Your dataset must score at least 100 in the formula below:

(NumberOfColumns * 4) * (NumberOfRows/100) >= 100
Examples:

A dataset with 5 columns and 10,000 rows would score: (5 * 4) * (10,000/100) = 2000
A dataset with 17 columns and 3,000 rows would score: (17 * 4) * (350/100) = 238
A dataset with 5 columns and 500 rows would score: (5 * 4) * (500/100) = 100

GRADING RUBRIC FOR FINAL PROJECT

Criteria	Points
House keeping: Your name and class (IST 719) clear on poster (can include a photo of you) Poster is minimum 24 x 36 to max 36 x 48. Poster can be portrait or landscape mode.	1
Presentation: practiced and delivered within instructor specified timeframe	1
Poster Story: Two sentences that provide overall context for the poster. In addition a meaningful poster title.	1
Motivation: 2 or 3 sentences that answer these questions: 1) who might be interested in this (audience); 2) why might they be interested? Ideally, this text leads to your questions (next item).	1

Questions: between 2 and 3 questions (with question mark) that can only be answered by relating two or more fields of data.	2
Data description text: 2 or 3 sentences that includes number of rows and columns, what the data is about and note subsetting, cleaning, and aggregations.	1
Sources: data source, R packages, R scripts, any templates, vector or raster images (small clip art, icons and small bits of code are excluded). Include link when possible.	1
Data descriptive plots: 2 to 4 single dimension plots that show the data distributions. Uses appropriate data encoding. Used R to make plot. These plots don't answer your questions. They provide insight into your dataset and help provide depth to your overall data story.	3
Key visualization: the main plot. Must be 2 or more dimensions showing a relationship that answers one of your questions. Uses appropriate data encoding. Use R to make plot. Should be in its own sub-heading area on poster with a relevant title.	3
Supporting Visualization(s): one or two additional multi dimension visualizations that answer your other questions. Uses appropriate data encoding. Used R to make plot. Should be in its own sub-heading area on poster with a relevant title.	2
Color: consistent, appropriate, appealing	2
Layout and use of space: good alignment, space for eye to rest, good navigation clues, not crowded	2
Image Quality: vector graphics or high resolution supporting images	2
Good overall design. Has WOW factor.	2
Files uploaded: Your original Adobe Illustrator file, **a pdf of the poster**, and all R script(s) that load data, do some data prep work, and make your plots. DO NOT INCLUDE THE DATA (but the poster should give me enough information that I could download it if I wanted). Files should be uploaded before live session 11.	1
Total	**25**

Specific ways I think about the poster as I grade it:
Could I reproduce your work? Could I find the data, select the same subset or do the same kinds of aggregations? What packages would I need? Did you have to do a lot of work to clean the data? If you used graphics or templates from the web, can I find and use them as well?

How is the overall effect? Does the poster support the "3 distances" (across the room appeal, middle distance topic overview, up close detail).

Does the layout and other visual elements support a visual hierarchy that enables quick navigation? Do you have clear sections for different areas (data description, areas for the two important plots, maybe an area for questions and motivation).

Consider, given that your data supports, using examples of some of the more advanced plots covered this semester ie. alluvial, network, leaf, spatial