

Guidelines for Infographic Project

Marketing Analytics
Professor César Zamudio

The Infographics Project is an important part of the Marketing Analytics class, because you will have to apply your knowledge of understanding business problems, data collection, visualization and even statistical inference, as well as your wit and skill in developing an infographic. The project will serve as useful practice for your professional life and will boost your analytics skills: dealing with a new analytics problem, developing data requirements and finding data, gathering it, visualizing it, and making editorial decisions concerning how to best present it.

The steps necessary for the project and useful guidelines are discussed in what follows, in the order in which you will deliver them.

Preliminaries

What is an infographic?

Because we will start thinking about the final project early on in the semester, you may not be entirely familiar with what an infographic is. Generally speaking, it is a set of visualizations with a strong storytelling component. This means that an infographic may be composed of not one, but multiple visualizations that, as a whole, tell a persuading or informative story about a given topic.

Who is the audience for an infographic? This is generally not a managerial audience, as for most of our activities and discussions during the semester. Instead, infographics are aimed to a general audience who must be informed or something, or who you may want to convince about something. With that said, recently, more companies have adopted an infographics approach to develop informative brochures, generally to persuade potential customers.

What are the characteristics of a good infographic? It should address an important topic, in a way that “sticks”: it should present compelling arguments while being interesting, memorable, and storytelling so as to maximize persuasion. Furthermore, if deployed on social media, a key goal of an infographic is to attain virality: one way to achieve this is to develop infographics which touch on key viral emotions such as humor, awe, or disgust, as years of research indicate.

What are good examples of editorial infographics?

Given that infographics are now quite prevalent in daily life, there are a plethora of great (and terrible) examples. Editorial infographics are those with a wider audience, that are

not used to promote a business but to persuade or inform about a given topic. Here are some examples:

- [Flowingdata's excellent repository of infographics](#), with explanations as to what makes them good (or bad).
- [Gareth Cook's "The Best American Infographics" book series](#). It is not free, but you may find it at your local library. Great selection of infographics, some of which are covered in class.
- [The Pudding](#), an online publication that promotes data-driven conversations. Interesting, although not exactly "infographics", but inspiring nonetheless!

What are good examples of **brand-centric infographics**?

Unlike traditional infographics, brand-centric infographics focus on your company. These would be used much like a brochure, to disseminate information about your company and why others might need your services. Here are some examples:

- Interfolio, a company that provides management of job applications, has a number of good infographics on their Dossier product: [this one](#) introduces you to the company and explains why letters of recommendation matter; [this other one](#) shows results from a survey of Dossier users.
- Import.io, a web data harvesting company, has an interesting [infographic on the real estate market](#), highlighting their thought process of creating the infographic.
- IBM showcases many of their products and services through infographics, for example, [Weather and NASCAR](#); [Shipping and Blockchain](#); [Anatomy of the Ultimate Fan Experience](#); [The Future of Identity](#); etc.
- SAS provides information on how to create infographics using their software, occasionally providing infographics of their own such as a [Star Trek infographic](#) and a [Twitter "stickman" viz.](#)
- Deloitte is a consulting company providing multiple infographics based on their findings, such as [Consumers in Control](#); [2016 Global Life Sciences outlook](#).
- Euromonitor, a marketing research services firm, provides a couple interesting infographics: [Navigating Marketing Research](#) (not data-driven, but interesting design); [How to Succeed in Emerging Markets](#).
- Yelp, a services review platform, provides insight into their business and how Yelp can help business succeed with their tool Yelp for Business, featuring infographics such as [this one](#) (not exactly an infographic but small pieces of one, focusing on making an argument tying Yelp browsing to purchases); as well as this [giant repository of Yelp infographics](#) which is quite instructive and informative.

What are some **FREE DIGITAL resources** I can use to learn more about infographics?

Since the course only briefly touches on the subject of infographics, feel free to peruse the following books available **FOR FREE and DIGITALLY at the VCU library website**

so you can access them whenever you want. [This link will take you to the available books](#). I suggest:

- Infographics for Dummies by Justin Beegel (a great resource, including this [free cheat sheet](#));
- Infographics: The Power of Visual Storytelling by Jason Lankow (with an interesting classification of infographics into brand-centric, for firms, and editorial infographics, for a larger audience);
- Cool Infographics by Randy Krum (a deep dive into many infographic styles, layouts, etc.)
- The Power of Infographics by Mark Smickilas (focusing on understanding your audience; and
- The Best Magazine Design Photography, Illustration, Infographics & Digital by the Society of Publication Designers (includes pages upon pages of excellent visual design, for inspiration). The [Amazon link](#) has many of these images publicly available;

Note that, aside from these digital resources, VCU has more books on-shelf not listed in the link provided above. Your local library (e.g., Richmond Public Library, Library of Virginia...) might also have some resources, so *check them out!*

What SOFTWARE can I use to make an infographic?

Since creating an infographic is a very creative endeavor, there is not one specific software you can use to create an infographic. Here are some software packages you might want to try:

General purpose software

- **Adobe Spark and Adobe Illustrator:** Spark can be used to create infographics using its [many infographics templates](#) (more [here](#)). Illustrator can be used to create original vector art. Because VCU is now an Adobe Creative Campus, we are now more engaged than ever with Adobe to provide their software to our students. The VCU student version of the Creative Suite (which includes Spark and Illustrator) [can be found here](#).
- **Vennage:** Web-based infographic creator. Features a very large number of [templates](#).
- **Canva:** An infographic creator very popular among the students, very similar to Vennage.
- **Easel.ly:** Yet another free option. Importantly, includes [a great set of tutorials on everything about infographics](#).
- **Snappa:** Another infographic creator, focusing on **novice users who have never used** photo editing software (e.g. Photoshop, Illustrator) before.
- **Infogram:** Infographics creator that is more data and chart-driven. Allows to embed actual data into your infographic and produce charts inside the web

platform itself. **Be sure to verify** that the data you use remains private in case you need to protect it.

Specialized software

- [MindTheGraph](#): Infographics creator focused on the sciences (e.g. medicine, physics). Useful if your infographic concerns those topics.
- [WordItOut](#): A word cloud generator which, importantly, includes a feature to manually filter irrelevant words.
- [Flourish](#): A powerful web visualization creator which includes many interesting viz. However, **the data in the free version is PUBLIC, so if you have data that you cannot/should not share (such as your employer's data) DO NOT use it.**
- [Biteable](#): Free video infographic creator, if you're interested in trying it.
- [GGPLOT2](#): A R library to create beautiful visualizations using R code. However, it requires some time to get used to it as it uses concepts from [The Grammar of Graphics](#).
- [Kartograph](#): A Python/JS geographic visualization library, if you want to learn it.
- [Blugraphic](#): A repository of free design templates to use as elements in your infographic.
- [Unsplash](#): A giant repository of free HD pictures to use as wallpaper or as elements/background(s) in your infographic.

Final Infographic Report I

The first Final Infographic Report requires you to sit and think about what topics you might want to explore this semester. It can be anything: past students have developed infographics based on Amazon mattress pricing; celebrity scandals; NCAA coach salaries; horror movie box office success; crime in the US before 2016; the growing number of animals in shelters; and many more. It's really up to you: **what excites YOU? What are YOU passionate or curious about?** In my experience, picking a topic you find interesting will positively influence the quality of your final work.

Your Final Infographic Report should contain the following.

- A list of **five** subjects you're interested in and **why you care about them** (be brief, a couple sentences is fine).
- An interesting **big question** you want to answer with the data.
- For each subject, list **at least one** possible source of data, **including a link to the website that hosts the data.**
- For each data source, briefly summarize it, focusing on the key variables and units of analysis (e.g., athletes, books, states...) it contains.

VERY IMPORTANT: what do we mean by “data”? You have to find a source that you can show has at least 50 unique observations for the topic you’re interested in. For example, a New York Times article reporting that there are three major causes of heart failure in the U.S. population is equivalent to **three** data points and **does not qualify**. Here are three examples of some **qualifying** topics and data sources a student might report:

- a. Stephanie’s passion is cup stacking, and she found a website from the WSSA containing a large amount of information, so she reports the following:

Topic	Why?	Big question	Data source	Data summary
Cup stacking	I am a professional cup stacking athlete	Which countries have the best and worst cup stackers?	WSSA records	Contains data on athletes and their performance by country

- b. Jamelle loves graphic novels, and he is interested in determining what the best and worst graphic novels have in common, so he reports the following:

Topic	Why?	Big question	Data source	Data summary
Graphic novels	I love graphic novels, especially those by DC Comics	<ul style="list-style-type: none"> • What are the best and worst evaluated graphic novels? • What do they have in common? 	Amazon Graphic Novels section	Data on novels, their overall rating, price, and reviews for every novel.

- c. Mona loves sneakers, and would like to know which are the most valuable sneakers of the past three years, so she reports the following:

Topic	Why?	Big question	Data source	Data summary
Sneaker prices	Sneakers are my passion and I am starting a collection!	<ul style="list-style-type: none"> • What are the most valuable sneakers of the past three years? • What makes them valuable? Brand, color, designer? 	StockX sneaker prices	Sneaker releases, characteristics, and pricing over time.

d. Qin is a student interested in healthcare issues and proposes the following:

Topic	Why?	Big question	Data source	Data summary
Health insurance and drug overdoses	I would like a job in healthcare and this project can help me understand important issues in the industry.	<ul style="list-style-type: none"> • Which states have the most drug overdoses? • Does lack of health insurance contribute to drug overdoses? 	<ol style="list-style-type: none"> 1. CDC Drug Overdose data (2014-2017) 2. Census Healthcare Coverage data 	<ul style="list-style-type: none"> • CDC data contains drug overdose by year and state. • Census contains healthcare coverage in same time frame.

Notice that in all the examples the students have found sources of data with **large tables, lists, or similar information that they can gather manually or using a scraper** (as we will learn later in the semester). Also, some projects seem easier and others more complicated, but, in general, **DO NOT UNDERESTIMATE** how difficult a project might be.

You might want to deliver a table with five topics as presented above, or be creative and use your own structure. However, please make sure you meet the minimum requirements (five ideas, with topic, reason, question, data source and brief summary).

Please deliver Final Infographic Report I via the Assignments section of Blackboard by the deadline specified in the syllabus.

- Use the format shown here or your Assignment will be penalized.
- Name your report “FirstNameLastName_FinalInfographicReportI” or your Assignment will be penalized.

Final Infographic Report II

Once you deliver Report I, you will have a set of ideas and datasets that you can explore this semester. **Do not start Report II before hearing from the professor: he will instruct you on which idea to pursue and which dataset to use.**

After the professor has let you know which idea you will tentatively pursue and which dataset you will use, the next step is to collect data. As discussed in class, **the time it will take you to collect and clean data may be considerable and should NEVER be underestimated.** Thus, the Final Infographic Report II requires you to show some preliminary evidence of your data collection, as well as a minimum analysis of the data you collected so far.

Your Final Infographic Report II should contain the following.

Data file

- A file containing the data you have collected so far. The data is expected to be provided in Excel format – however, advanced students might provide a Tableau TWBX file, which is fine as well. If you wish to submit other data formats, consult with the professor first.
 - **IMPORTANT: If you downloaded data from the Web or using WebHarvy, FIRST SAVE IT AS A CSV FILE. Then, open it, and SAVE AS... EXCEL FILE. The CSV version is for your personal backup, and the Excel file is what you will deliver.**
- Variable names that satisfy the following requirements. These requirements will help you develop a good sense of “etiquette” when constructing your datasets so that other members of an analytics team (in this case, the professor) can understand your work:
 - Each variable name should be reasonably legible. For example, “WTPSegment” probably means willingness to pay of certain segments. Names such as “bk”, “sd12”, “ar_34”, “variab”, or similar, should not be included.
 - No variable name should start with a number. For instance, you may want to call a variable “Insurance2” or “S2_ScaleItem”. But do not write “2Insurance” or “2S_ScaleItem”.
 - Variables can be named using lowercase or uppercase as you see fit.
 - Variable names should not be excessively long – think a maximum of 20 characters or so.
- Relevant IDs that are correctly identified and marked as such. For instance, you may indicate a consumer ID as “ConsumerID”, “ID_Consumer” or “CID”; transactions might be called “Trans_ID”; basketball games “IDGame” and so forth. If your dataset has multiple levels, please indicate the corresponding ID for each. **If your levels have IDs that can be represented by text (e.g., State names, country names) you can use those as IDs.**
- **If your dataset is an Excel file**, consider coloring your columns and variable headings in a way that is pleasant and indicates different variable groups. Fig. 1 presents an example. Feel free to create a style that suits your own needs. The goal is to make the structure of the data easy to understand and the major variable groups easy to remember.

Fig. 1. Example of colored variable headings and columns in an Excel File

EN	EO	EP	EQ	ER	ES	ET	EU	EV
ATTCHECK2	KnowCReports	CREPORTSKNOW	TrustCReports	Demo_Age	Demo_Gender	MALE	FEMALE	OTHER_GENDER
4	2	0		33	1	1	0	0
4	1	1	5	27	1	1	0	0
4	1	1	4	23	1	1	0	0
4	1	1	5	26	1	1	0	0
4	2	0		27	2	0	1	0
4	1	1	7	31	1	1	0	0
4	2	0		25	2	0	1	0
1	1	1	5	26	2	0	1	0
4	1	1	7	26	1	1	0	0
4	2	0		66	1	1	0	0
4	1	1	5	47	1	1	0	0
4	2	0		40	2	0	1	0
4	2	0		20	1	1	0	0
4	2	0		20	1	1	0	0

Data dictionary

The data dictionary can be delivered as a Word or Excel file (if your data is an Excel file itself, the data dictionary can be a separate Sheet) and must contain at a minimum, **for each variable** in the data:

- The name of the variable.
- The level the variable belongs to. For example, five variables may pertain to State level and four to Factory level.
- (Optional) If some of your variables belong to identifiable groups, which group each variable belongs to. For example, ten variables might be about Attitude towards Apple and five about Demographics. You are free to define these groups yourself and doing so is optional.
- A short description of what the variable is, so that other analysts can easily figure this out. If the variable is nominal or ordinal try to describe, in general terms, what the categories are. **Be sure to include the NUMBER of total categories.**
- The measurement level (NOIR, Unstructured, ID or Date)
- The units of measurement (e.g., dollars, thousands of dollars, 5-point Likert scale, 10-point rating scale, 5-point online review...)
- Optionally, any additional remarks about the variable that a different analyst might need to know when using your dataset. For instance, a certain scale might come from a particular research paper, or a variable might have been collected specifically for a certain goal.

Table 1 presents one example of how a good data dictionary might look like. **FOLLOW THIS EXAMPLE. (Next Page)**

Table 1: Example of data dictionary for a hamburger restaurant

Level	Group	Var. name	Description	Measurement	Units	Remarks
Consumer	-	CID	Consumer ID	ID	Numeric ID	
	-	Fav_Rest	Consumers' favorite hamburger restaurant (among five: BK, McD, Wendy's, Hardee's, Cookout)	Nominal	-	
	-	Rating	Rating of consumers' last trip to our restaurant	Interval	10-star scale	
	Attitude towards hamburgers	S1_Like	Item 1: I like hamburgers	Interval	5-point agreement scale	Source for the scale is the work of Hungry et al. (2015)
		S2_Flavor	Item 2: Hamburgers have great flavor	Interval		
S3_Convenient		Item 3: Hamburgers are convenient	Interval			
State	-	SID	State ID	ID	Text ID	
	-	Population	Population in state as of 2017	Ratio	Millions	From Census
	-	Weather	Average weather in state as of 2017	Ratio	Fahrenheit	From AccuWeather

Note how, in this table, the hamburger restaurant names are abbreviated informally (“McD”, “BK”) for brevity. You can do so for any content of the dictionary as long as it is obvious enough like in this case.

Descriptives

The descriptives can be delivered as a Word or Excel file (if your data is an Excel file itself, the data dictionary can be a separate Sheet). For every variable in the data you must provide the following:

- **For continuous (measures) data:**
 - The minimum;
 - The maximum;
 - The average;
 - The standard deviation;
- **For discrete (dimensions) data:**
 - You must identify the categories contained in the data. For instance, if you have an ordinal variable of Olympic medals, you'd have three categories: Bronze, Silver, Gold; you might have five cities in a given variable; or you may have a Yes/No dummy. For each category:
 - The Percentage of observations within the category (for example, if you recorded 300 Yes and 700 No, you would include 30% Yes and 70% No)
 - If you have a very large amount of categories (more than 7) then report the most common 5 categories.

- **You do not need to provide descriptives for ID or Date variables.** However, do provide the following:
 - Number of total observations in the data
 - **For each level in your data**, the total number of observations. For example: “one million reviews from 375 consumers from 8 U.S. states about 4 brands of coffee”.
- **Optionally**, provide the corresponding level if it makes the table easier to read.

Table 2 presents an example of an appropriate descriptives table.

Table 2. Descriptives table for a hamburger restaurant, following Table 1

Level	Variable	Avg. or Count	Pct.	Min	Max	Std. Dev.
Consumer	Fav_Rest	86	100	-	-	-
	<i>Burger King</i>	25	29.07	-	-	-
	<i>McDonald's</i>	45	52.33	-	-	-
	<i>Wendy's</i>	1	1.16	-	-	-
	<i>Hardee's</i>	12	13.95	-	-	-
	<i>Cookout</i>	3	3.49	-	-	-
	Rating	8.75	-	1	10	2.25
	S1_Like	4.21	-	2	5	0.71
	S2_Flavor	4.78	-	2	5	0.12
	S3_Convenient	3.90	-	1	5	2.78
State	Population	25.12	-	3.57	39.55	12.21
	Weather	75.30	-	53.09	105	21.12

Note how, in this example:

- The average (for continuous variables) and the count (for nominal variables) are collapsed in a single column. This is because those two metrics are the most important in general.
- The nominal variable is collapsed in a pleasant way, by including the five categories below the variable name, appropriately identified using italics and indentation, and presenting the breakdown of responses for each one. Note that the counts sum to 86 and the percentages sum to 100. **This has to be the case for each of your nominal or ordinal level variables.**
- Given their nature, for continuous variables, the percentages are omitted; for nominal variables, the min, max, and standard deviation are omitted.

Your Final Infographic Report II must contain your data, data dictionary and descriptives. **Please deliver Final Infographic Report II via the Assignments section of Blackboard by the deadline specified in the syllabus.**

- Use the formats in Figure 1, Table 1 and Table 2 as a guide of how a professional report should look like. Assignments that do not pay attention to formatting details will be penalized.
- Name your report “FirstNameLastName_FinalInfographicReport2” or your Assignment will be penalized.

If you have NOT completed the Report by the deadline, you must provide evidence of sufficient progress and can request up to ONE-week extension. No additional extensions will be permitted.

Final Infographic Report III

After completing your Report III, you have:

- An idea you chose to pursue; and
- A clean dataset you proposed, with a data dictionary;

For the Final Report III, simply deliver a draft of your final infographic in whichever format you are developing it (most likely PDF, an image file, or even PPT). This is all you need to deliver. I will use this draft to provide feedback in preparation for your final deliverable. Good luck! What I expect/useful tips:

- Have an interesting title.
- State your name and whatever information about you you'd like to share.
- Have an introductory paragraph describing the main goal of your project – what inspired you? Why are you curious about this problem? Try to start a narrative here. However, be as brief as possible.
- Include the **SOURCES** you used. For instance, if you scraped data from BeerAdvocate.com, be sure to cite them. Maybe even include their logo as part of your Infographic.
- Supplement your Infographic with images and other appealing content. For instance, if making an Infographic about movies, why not include some poster art? If your Infographic is about sneakers, why not show a couple sneakers?
- When placing images in your Infographic, make sure **you do not DEFORM** the image if you resize it. To this effect, when resizing an image, hold SHIFT (or CTRL, or CMD, depending on your computer and software – experiment!) while resizing the image with your mouse. This will preserve the image's “aspect ratio” so it does not look deformed.

ONLY REQUIREMENT: You cannot have a PPT with multiple slides. Your Infographic must be ONE SINGLE image.

Final Infographic Deliverable

Your Final deliverable should include:

- Your infographic file (PDF, PPT, etc.)
 - Again, if you are delivering on PPT, you cannot have a PPT with multiple slides. Use one single image.
- The dataset you used to construct the infographic

Make sure that the final infographic has your name, and that it adequately references the source data/organization/etc. from which you obtained the data.

A file in Canvas contains some examples of good infographics you can refer to. As a rough guideline for your grade:

- Infographics that do not have at least three **data-driven** charts (i.e., using the actual data you gathered) will likely obtain no more than 50% of the final infographic points.
- Infographics that rely on very small datasets will likely lose about 30% of the final infographic points.
- Infographics that do not have a professional appearance or exhibit multiple typos will be penalized depending on how prevalent this problem is.