

DOCUMENT EVERYTHING

Maureen O'Donnell

Dr. Arnholt Senior Seminar

odonnellm@appstate.edu

February 23, 2014

Overview

1 What To Document?

- Where did you get the data?
- How to Store the Data?

2 Our Project

3 In Summary

Where did you get the data?

The location of where your data is from is very important to document. In the data frame used in the NC Voter project had 4 sources of data. As you will soon see we did not document the data well.

When documenting where your data is from, be sure to include how you got it. Did you have to click on multiple tabs within the page? Was it a direct url? How often is the data changed?

Where Our Data Was From

- NC Voter Info Site
 - Easiest site to use of the bunch
- US Census Bureau
 - Not a good idea to use this site. In fact DO NOT use it. It is miserable to search through and very hard to find a simple file.
- Two Church sites
 - Not as bad to use. However we did not document the location within the site so finding the data has proved difficult.
- North Carolina Education site
 - Also not too bad to use. Again we did not document where in the site the data was.

URL's of the sites we used

NC Voter

<http://www.ncsbe.gov/>

Census

<http://factfinder2.census.gov>

Churches

<http://www.thearda.com>

Education

<http://collegestats.org/colleges/north-carolina>

How to Store the Data

- Pick one place for all group members to store the data
- Don't use 1 billion different locations to store the data, like we did.
- Label any changes you make so all the group can see.
- Date all changes

Screen Shot 1

Microsoft Excel																						
Home										Insert	Page Layout	Formulas	Review	View	Developer	Add-Ins						
Calibri										11	A	A	A	A	A	A	A	A	A	A	A	
Font										B	C	D	E	F	G	H	I	J	K	L	M	
Paragraph										N	O	P	Q	R	S	T	U	V	W	X	Y	
Clipboard										Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
Clipboard										AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY
Clipboard										AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL
Clipboard										BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY
Clipboard										BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL
Clipboard										CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY
Clipboard										CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ
Clipboard										DK	DL	DM	DN	DO	DP	DQ	DR	DS	DT	DU	DV	DW
Clipboard										DX	DY	DZ	EA	EB	EC	ED	EE	EF	EG	EH	EI	EJ
Clipboard										EK	EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU	EV	EW
Clipboard										EX	EY	EZ	FA	FB	FC	FD	FE	FF	FG	FH	FI	FJ
Clipboard										FK	FL	FM	FN	FO	FP	FQ	FR	FS	FT	FU	FV	FW
Clipboard										FX	FY	FZ	GA	GB	GC	GD	GE	GF	GG	GH	GI	GJ
Clipboard										GK	GL	GM	GN	GO	GP	GQ	GR	GS	GT	GU	GV	GW
Clipboard										GX	GY	GZ	HA	HB	HC	HD	HE	HF	HG	HH	HI	HJ
Clipboard										HK	HL	HM	HN	HO	HP	HQ	HR	HS	HT	HU	HV	HW
Clipboard										HX	HY	HZ	IA	IB	IC	ID	IE	IF	IG	IH	II	IJ
Clipboard										IK	IL	IM	IN	IO	IP	IQ	IR	IS	IT	IU	IV	IW
Clipboard										IX	IY	IZ	JA	JB	JC	JD	JE	JF	JG	JH	JI	JJ
Clipboard										JK	JL	JM	JN	JO	JP	JQ	JR	JS	JT	JU	JV	JW
Clipboard										JX	JY	JZ	KA	KB	KC	KD	KE	KF	KG	KH	KI	KJ
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW
Clipboard										KK												

Figure : A portion of our data

Screen Shot 2

Clipboard	Font	Alignment	Number	Styles	Cells	Editing
H34						
	A	B	C	D		
1						
2	Initial Name	New Name	Explanation	Where is the data from?		
3	PERCENTFOR	pollfor	the percent of voters that voted for amendment 1			
4	PercentVoterTurnout	turnout	the percent of registered voters that voted		These data come from	
5	Pres_2008_Obama_Percent	obama08	the percent of voters that voted for Obama in the 2008 election		http://www.ncsbn.gov/votest08.asp?id=43	
6	Pres_2008_McCain_Percent	mccain08	the percent of voters that voted for McCain in the 2008 election			
7	TOTPOP2010	pop2010	the total population of the county in 2010			
8	MedianAge	medage	the median age of the county population			
9	Divorced	divorced	was this data found the same time as the total population			
10	PercentRural	pcrural	The percent of the population that lives in rural areas of the county			
11	TotalAdvancedDegreesBAorabove	edu_baorup	The total population of people with an advanced degree of a BA			
12	MedianHouseholdIncomepercapita2010	medinc	the median household income adjusted for inflation			
13	EnrollmentCollegeGrad	col_enroll	the number of people enrolled in colleges in the county		These data come from	
14	Age18to24	age18-24	the number of people between the ages of 18 and 24 in the county		http://actfinder2.census.gov/ageandsex/ageandsexresults.shtml?ref=shs#	
15	Age25to59	age25-59	the number of people between the ages of 25 and 59 in the county			
16	Age60plus	age60up	the number of people over 60			
17	TotalEnrollmentCollegeStudents	students	the number of students enrolled in the colleges in the county			
18	WhiteNumber	white	the number of white people in the county			
19	BlackNumber	black	the number of black people in the county			
20	PopulationDensity	popden	the population density of the county			
21	NumberofColleges	colleges	the number of colleges in a given county		http://collegedata.org/collegedata-carolina	
22	EvangelicalRate	evangelical			http://www.thearda.com	
23	Churches	churches	the number of churches in the county			
24						
25		collden	The density of colleges in a given county			
26		MaleBachelor	The number of Male with a Bachelor's degree			
27		MaleMaster	The number of Male with a Master's degree			
28		MaleProfessional	The number of Male with a Professional degree			
29		MaleDoctorate	The number of Male with a Doctorate			
30		FemaleBachelor	The number of Females with a Bachelor's degree		I am not 100% sure where these data values came from	
31		FemaleMaster	The number of Females with a Master's degree		I think they are from the census website as well.	
32		FemaleProfessional	The number of Females with a Professional degree			
33		FemaleDoctorate	The number of Females with a Doctorate degree			

Figure : Our variable table

Screen Shot 3

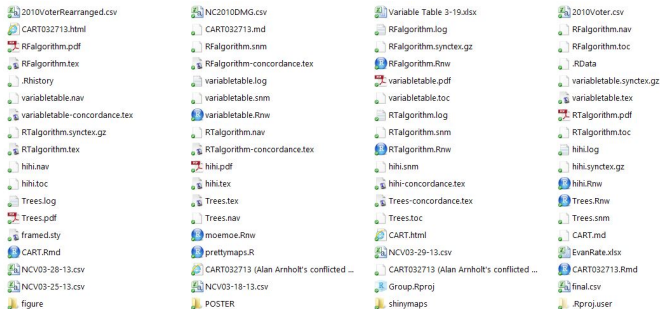


Figure : One of our messy folders

What to take away

Large data sets can seem difficult to manage but with appropriate storage techniques it can be quite easy.

Once all the data is stored the resulting R package will use what you've done to make is user friendly to anyone wishing to do analysis with it.

The End