

# Math 107

Introduction to Data

# Why Statistics?

- Statistics is all about **thinking with and about data**
- Data are everywhere! Regardless of your field, interests, lifestyle, etc., you will almost definitely have to make decisions based on data, or evaluate decisions someone else has made based on data

# The Basics

# Data

Data are a set of measurements taken on a set of individuals

We often see data presented as a dataset, comprised of variables measured on cases

# Campaign Expenditures

- FEC link and codebook linked from course webpage
- What are the cases?
- What are the variables?

Spender Name	Candidate Name	Support/ Oppose	State	Dist	Office	Party	Filing Date	Expend Date	Expend Amount	Purpose
2016 COMMITTEE; THE	Carson, Ben	Support		00	President		11/04/2015		\$4,000.00	Skywalk Panels with Brochure Distribution
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President		01/20/2016	07/14/2015	\$3,725.00	Television Ad Placement
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President		01/20/2016	07/15/2015	\$12,016.86	Television Ad Production
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President	Republican	08/05/2015	07/15/2015	\$324,795.00	Television Ad Placement
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President	Republican	08/05/2015	07/14/2015	\$3,725.00	Television Ad Placement
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President	Republican	08/05/2015	07/15/2015	\$12,016.86	Television Ad Production
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President		01/20/2016	07/15/2015	\$376,907.00	Television Ad Placement
AMERICA'S RENEWABLE FUTURE INC	CRUZ, RAFAEL EDWARD 'TED'	Oppose	IA	00	President	Republican	03/11/2016	01/25/2016	\$9,000.00	Literature

# Categorical vs. Quantitative

A **categorical** variable divides cases into groups, and is recorded as a label

A **quantitative** variable measures a numerical quantity

- recorded and used as a number
- arithmetic makes sense

# Campaign Expenditures

- Identify each variable as either categorical or quantitative

Spender Name	Candidate Name	Support/ Oppose	State	Dist	Office	Party	Filing Date	Expend Date	Expend Amount	Purpose
2016 COMMITTEE; THE	Carson, Ben	Support		00	President		11/04/2015		\$4,000.00	Skywalk Panels with Brochure Distribution
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President		01/20/2016	07/14/2015	\$3,725.00	Television Ad Placement
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President		01/20/2016	07/15/2015	\$12,016.86	Television Ad Production
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President	Republican	08/05/2015	07/15/2015	\$324,795.00	Television Ad Placement
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President	Republican	08/05/2015	07/14/2015	\$3,725.00	Television Ad Placement
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President	Republican	08/05/2015	07/15/2015	\$12,016.86	Television Ad Production
AMERICA NEXT INC	Jindal, Bobby	Support	IA	00	President		01/20/2016	07/15/2015	\$376,907.00	Television Ad Placement
AMERICA'S RENEWABLE FUTURE INC	CRUZ, RAFAEL EDWARD 'TED'	Oppose	IA	00	President	Republican	03/11/2016	01/25/2016	\$9,000.00	Literature

# Explanatory vs. Response

When we are investigating **relationships** between variables, we often think about two types of variables.

- The result/outcome of interest is called the **response variables**
- The variable(s) used to understand/predict the response variable are called **explanatory variables**



# **Tidy Data**

“Happy families are all alike; every unhappy family is unhappy in its own way.”

–Leo Tolstoy

# Tidy Data

- Each variable forms a column
- Each case forms a row
- No footnotes
- No units