

# Data Analysis and Statistical Inference

## Introduction

Statistics 101

Dr. Çetinkaya-Rundel

January 7, 2015

- 1 General info
- 2 Example: Baby names
- 3 Example: Geotagged data
- 4 Why study statistics?

- ▶ Professor: Dr. Mine Çetinkaya-Rundel - [mine@stat.duke.edu](mailto:mine@stat.duke.edu)
- ▶ TAs:
  - TBA

- ▶ OpenIntro Statistics, 2nd Edition
- ▶ i>clicker2 - See Google Doc for a list of students selling used clickers (link emailed)
- ▶ (optional) Calculator

*<http://bit.ly/sta101sp15>*

Component	Weight
Attendance & participation + peer evaluation	7.5%
Problem sets	10%
Labs	10%
Readiness assessments	10%
Performance assessments	2.5%
Project 1	5%
Project 2	10%
Midterm 1	10%
Midterm 2	10%
Final	25%

- ▶ Grades curved at the end of the course after overall averages have been calculated.
  - average of 90-100 guaranteed A-
  - average of 80-90 guaranteed B-
  - average of 70-80 guaranteed C-
- ▶ The more evidence there is that the class has mastered the material, the more generous the curve will be.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.



- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
- ▶ Use statistical software to summarize data numerically and visually, and to perform data analysis.
- ▶ Have a conceptual understanding of the unified nature of statistical inference.
- ▶ Apply estimation and testing methods to analyze single variables or the relationship between two variables in order to understand natural phenomena and make data-based decisions.
- ▶ Model numerical response variables using a single or multiple explanatory variables.
- ▶ Interpret results correctly, effectively, and in context without relying on statistical jargon.
- ▶ Critique data-based claims and evaluate data-based decisions.
- ▶ Complete two research projects: one that focuses on statistical inference and one that focuses on modeling.

- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.

- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.
- ▶ *Unit 2 - Probability and distributions:* The basics of probability and chance processes, Bayesian perspective in statistical inference, the normal distribution.



- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.
- ▶ *Unit 2 - Probability and distributions:* The basics of probability and chance processes, Bayesian perspective in statistical inference, the normal distribution.
- ▶ *Unit 3 - Framework for inference:* Central Limit Theorem and sampling distributions

- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.
- ▶ *Unit 2 - Probability and distributions:* The basics of probability and chance processes, Bayesian perspective in statistical inference, the normal distribution.
- ▶ *Unit 3 - Framework for inference:* Central Limit Theorem and sampling distributions
- ▶ *Unit 4 - Statistical inference for numerical variables*

- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.
- ▶ *Unit 2 - Probability and distributions:* The basics of probability and chance processes, Bayesian perspective in statistical inference, the normal distribution.
- ▶ *Unit 3 - Framework for inference:* Central Limit Theorem and sampling distributions
- ▶ *Unit 4 - Statistical inference for numerical variables*
- ▶ *Unit 5 - Statistical inference for categorical variables*

- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.
- ▶ *Unit 2 - Probability and distributions:* The basics of probability and chance processes, Bayesian perspective in statistical inference, the normal distribution.
- ▶ *Unit 3 - Framework for inference:* Central Limit Theorem and sampling distributions
- ▶ *Unit 4 - Statistical inference for numerical variables*
- ▶ *Unit 5 - Statistical inference for categorical variables*
- ▶ *Unit 6 - Simple linear regression:* Bivariate correlation and causality, introduction to modeling

- ▶ *Unit 1 - Introduction to data:* Observational studies and non-causal inference, principles of experimental design and causal inference, exploratory data analysis: description, summary and visualization, introduction to statistical inference.
- ▶ *Unit 2 - Probability and distributions:* The basics of probability and chance processes, Bayesian perspective in statistical inference, the normal distribution.
- ▶ *Unit 3 - Framework for inference:* Central Limit Theorem and sampling distributions
- ▶ *Unit 4 - Statistical inference for numerical variables*
- ▶ *Unit 5 - Statistical inference for categorical variables*
- ▶ *Unit 6 - Simple linear regression:* Bivariate correlation and causality, introduction to modeling
- ▶ *Unit 7 - Multiple linear regression:* More advanced modeling with multiple predictors

- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit
- ▶ Prior to beginning the unit, complete the readings and familiarize yourselves with the learning objectives
- ▶ Begin a new unit with a readiness assessment: individual, then team
- ▶ Class time: split between lecture, discussion/application, and lab
- ▶ Complement your learning with problem sets
- ▶ Wrap up a unit with a performance assessment

- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit
- ▶ Prior to beginning the unit, complete the readings and familiarize yourselves with the learning objectives
- ▶ Begin a new unit with a readiness assessment: individual, then team
- ▶ Class time: split between lecture, discussion/application, and lab
- ▶ Complement your learning with problem sets
- ▶ Wrap up a unit with a performance assessment

- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit
- ▶ Prior to beginning the unit, complete the readings and familiarize yourselves with the learning objectives
- ▶ Begin a new unit with a readiness assessment: individual, then team
- ▶ Class time: split between lecture, discussion/application, and lab
- ▶ Complement your learning with problem sets
- ▶ Wrap up a unit with a performance assessment



- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit
- ▶ Prior to beginning the unit, complete the readings and familiarize yourselves with the learning objectives
- ▶ Begin a new unit with a readiness assessment: individual, then team
- ▶ Class time: split between lecture, discussion/application, and lab
- ▶ Complement your learning with problem sets
- ▶ Wrap up a unit with a performance assessment

- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit
- ▶ Prior to beginning the unit, complete the readings and familiarize yourselves with the learning objectives
- ▶ Begin a new unit with a readiness assessment: individual, then team
- ▶ Class time: split between lecture, discussion/application, and lab
- ▶ Complement your learning with problem sets
- ▶ Wrap up a unit with a performance assessment

- ▶ Set of learning objectives and required and suggested readings, videos, etc. for each unit
- ▶ Prior to beginning the unit, complete the readings and familiarize yourselves with the learning objectives
- ▶ Begin a new unit with a readiness assessment: individual, then team
- ▶ Class time: split between lecture, discussion/application, and lab
- ▶ Complement your learning with problem sets
- ▶ Wrap up a unit with a performance assessment

- ▶ Highly functional teams of learners based on survey and pre-test
- ▶ Team members first point of contact
- ▶ Application exercises, team readiness assessments, projects
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work
- ▶ Peer evaluations to ensure that all team members contribute to the success of the group and to address any potential issues early on
- ▶ If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA's attention ASAP (don't wait till things get worse)

- ▶ Highly functional teams of learners based on survey and pre-test
- ▶ Team members first point of contact
- ▶ Application exercises, team readiness assessments, projects
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work
- ▶ Peer evaluations to ensure that all team members contribute to the success of the group and to address any potential issues early on
- ▶ If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA's attention ASAP (don't wait till things get worse)

- ▶ Highly functional teams of learners based on survey and pre-test
- ▶ Team members first point of contact
- ▶ Application exercises, team readiness assessments, projects
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work
- ▶ Peer evaluations to ensure that all team members contribute to the success of the group and to address any potential issues early on
- ▶ If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA's attention ASAP (don't wait till things get worse)

- ▶ Highly functional teams of learners based on survey and pre-test
- ▶ Team members first point of contact
- ▶ Application exercises, team readiness assessments, projects
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work
- ▶ Peer evaluations to ensure that all team members contribute to the success of the group and to address any potential issues early on
- ▶ If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA's attention ASAP (don't wait till things get worse)

- ▶ Highly functional teams of learners based on survey and pre-test
- ▶ Team members first point of contact
- ▶ Application exercises, team readiness assessments, projects
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work
- ▶ Peer evaluations to ensure that all team members contribute to the success of the group and to address any potential issues early on
- ▶ If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA's attention ASAP (don't wait till things get worse)



- ▶ Highly functional teams of learners based on survey and pre-test
- ▶ Team members first point of contact
- ▶ Application exercises, team readiness assessments, projects
- ▶ Study together, but anything that is not explicitly a team assignment must be your own work
- ▶ Peer evaluations to ensure that all team members contribute to the success of the group and to address any potential issues early on
- ▶ If you feel that there are issues within your team, you are encouraged to discuss it with your team members and to bring it to my or your TA's attention ASAP (don't wait till things get worse)

*Objective:* Two-way communication and instant feedback

- ▶ Readiness assessments (graded for accuracy)
- ▶ Questions throughout lecture (graded for participation)
  - to get credit for the day you must respond to at least 75% of the questions
  - up to three unexcused late arrivals or absences will not affect your clicker grade
- ▶ Register your clicker
  - <https://www1.iclicker.com/register-clicker> (Student ID = Net ID)
  - grading starts Mon, Jan 26

*Objective:* Make you an active participant and help me pace the class.

- ▶ Attendance and participation during class, as well as your activity on Piazza make up a non-insignificant portion of your grade in this class
- ▶ Might sometimes call on you during the class discussion, however it is your responsibility to be an active participant without being called on

*Objective:* Help you develop a more in-depth understanding of the material and help you prepare for exams and projects.

- ▶ Questions from the textbook
- ▶ Show all your work to receive credit, type your answers in the text box on Sakai (and attach any plots/images as separate files, properly named) or attach a PDF of your answers (no other file types will be accepted)
- ▶ Welcomed and encouraged to work with others, but turn in your own work
  - If you copy someone else's work, both parties will receive a 0 for the problem set grade as well as being reported to the Office of Student Conduct
- ▶ Lowest score will be dropped
- ▶ No make-ups
- ▶ Excused absences (e.g. STINF) do not excuse homework

*Objective:* Give you hands on experience with data analysis using statistical software and provide you with tools for the projects.

- ▶ Work in teams: author / discussants
- ▶ Must be present in lab session to get credit
- ▶ Lowest lab score will be dropped

*Objective:* Give you hands on experience with data analysis using statistical software and provide you with tools for the projects.

- ▶ Work in teams: author / discussants
- ▶ Must be present in lab session to get credit
- ▶ Lowest lab score will be dropped

### Activity: Get started with R/RStudio

- ▶ Go to the course website, <http://bit.ly/sta101sp15>, and click on the RStudio link on the top right
- ▶ Log in using your Net ID and password
- ▶ In the Console, generate a random number between 1 and 5, and introduce yourself to that many people sitting around you:  
`sample(1:5, size = 1)`

*Objective:* Encourage you to watch the videos and complete the reading assignment prior to coming to class and evaluate your conceptual understanding of the learning objectives.

- ▶ 10 multiple choice questions, at the beginning of a unit
- ▶ Conceptual questions addressing the learning objectives of the new unit, assessing familiarity and reasoning, not mastery
- ▶ Take the individual RA using clickers, then re-take in teams
- ▶ Individual RA score  $\frac{3}{4}$  of grade, team RA score  $\frac{1}{4}$  & your input during the team portion will factor into your participation grade
- ▶ Lowest score will be dropped

*Objective:* Evaluate your mastery of the material by the end of a unit and give you instant feedback on your performance.

- ▶ 10 multiple choice questions, at the end of a unit
- ▶ Taken individually on Sakai
- ▶ Lowest score will be dropped



*Objective:* Give you independent applied research experience using real data and statistical methods.

- ▶ Project 1: For a parameter of interest to you, you will describe the relevant data, compute a confidence interval and conduct a hypothesis test, and summarize your findings in a written, fully reproducible, data analysis report
- ▶ Project 2: You will use all (relevant) techniques learned in this class to analyze a dataset provided by me. You will share your results in a poster session
- ▶ Must complete both projects and score at least 30% of the points on each project in order to pass this class

Midterm 1	Wed, Feb 18
Midterm 2	Wed, Mar 25
Final	Sat, May 2 (2-5pm)

- ▶ Exam dates cannot be changed, no make-up exams will be given
- ▶ If you cannot take the exams on these dates you should drop this class.
- ▶ Calculator + cheat sheet allowed

- ▶ I will regularly send announcements by email, so make sure to check your email daily
- ▶ While email is the quickest way to reach me outside of class, sometimes it may be more efficient to answer most statistical questions "in person" so make use of OH
- ▶ Any non-personal questions related to the material covered in class, problem sets, labs, projects, etc. should be posted on Piazza forum
- ▶ Before posting a new question please make sure to check if your question has already been answered, and answer others' questions

Students with disabilities who believe they may need accommodations in this class are encouraged to contact the [Student Disability Access Office](#) at (919) 668-1267 as soon as possible to better ensure that such accommodations can be made.

<http://www.access.duke.edu/students/requesting/index.php>

- ▶ Late work policy for problem sets and labs reports:
  - next day: lose 30% of points (within 24 hours of due date)
  - later than next day: lose all points
- ▶ Late work policy for projects: 10% off for each day late

Regrade requests must be made *within 3 days* of when the assignment is returned, and must be submitted in writing.

- ▶ These will be honored if points were tallied incorrectly, or if you feel your answer is correct but it was marked wrong
- ▶ No regrade will be made to alter the number of points deducted for a mistake
- ▶ There will be no grade changes after the final exam

- ▶ No make-up for attendance, individual and team readiness assessments, labs, problem sets, projects, or exams
- ▶ If the midterm exam must be missed due to a documented medical excuse, absence must be officially excused *in advance*, in which case the missing exam score will be imputed using the final exam score
- ▶ The final exam must be taken at the stated time
- ▶ If you can't take the exams on these dates you should drop this class
- ▶ You must take the final exam and turn in the projects in order to pass this course

- ▶ Clickers may not be shared, and the clicker registered to a person may only be used by that person, failure to abide by this will result in a 0 clicker grade for everyone involved
- ▶ Use of disallowed materials (textbook, class notes, web references, any form of communication with classmates or other persons, etc.) during exams will not be tolerated



Any form of academic dishonesty will result in an immediate 0 on the given assignment and will be reported to the Office of Student Conduct. Additional penalties may also be assessed if deemed appropriate. If you have any questions about whether something is or is not allowed, ask me beforehand.

Some examples:

- ▶ Use of disallowed materials (including any form of communication with classmates or accessing the web) during exams and readiness assessments.
- ▶ Plagiarism of any kind.
- ▶ Use of outside answer keys or solution manuals for the homework.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

- ▶ Complete the reading before a new unit begins, and then review again after the unit is over.
- ▶ Be an active participant during lectures and labs.
- ▶ Ask questions - during class or office hours, or by email. Ask me and your classmates.
- ▶ Do the problem sets - start early and make sure you attempt and understand all questions.
- ▶ Start your projects early and allow adequate time to complete them.
- ▶ Give yourself plenty of time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- ▶ Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.



- ▶ Download or purchase the textbook
- ▶ Obtain and register your clicker
- ▶ Complete the pretest and survey
- ▶ Read the syllabus and let me know if you have any questions
- ▶ Start reviewing the resources for Unit 1 - readiness assessment on Monday (not graded, for practice)

- 1 General info
- 2 Example: Baby names
- 3 Example: Geotagged data
- 4 Why study statistics?

- ▶ Each year the Social Security Administration collects and releases data on the how many babies are given a certain name.
- ▶ They released these data for years 1880 onwards for each gender.
- ▶ For privacy reasons they restrict the list of names to those with at least 5 occurrences.



## Just Arrived!

Social Security's Top 10 Baby Names for 2013!

Rank	Male name	Female name
1	Noah	Sophia
2	Liam	Emma
3	Jacob	Olivia
4	Mason	Isabella
5	William	Ava
6	Ethan	Mia
7	Michael	Emily
8	Alexander	Abigail
9	Jayden	Madison
10	Daniel	Elizabeth

<http://www.babynamewizard.com/voyager>

## NameVoyager: Explore baby names and name trends letter by letter

Looking for the perfect baby name? **Sign up for free** to receive access to our expert tools!

Baby Name >

☐ Both ☐ Boys ☒ Girls

Current rank: boys 

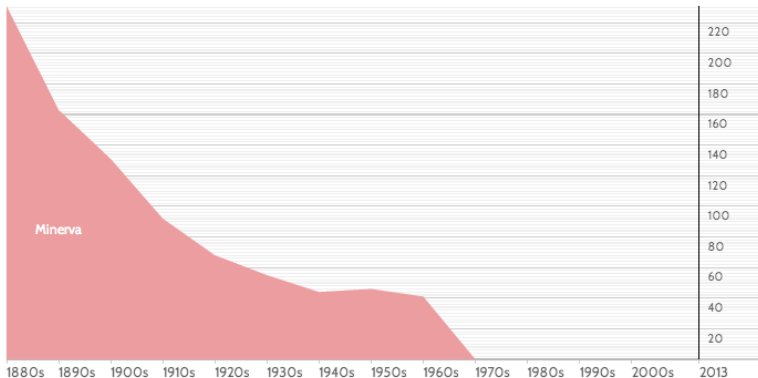
1000	500	100	25	1
------	-----	-----	----	---

  
girls 

1000	500	100	25	1
------	-----	-----	----	---

Names starting with 'MINE' per million babies

per million births

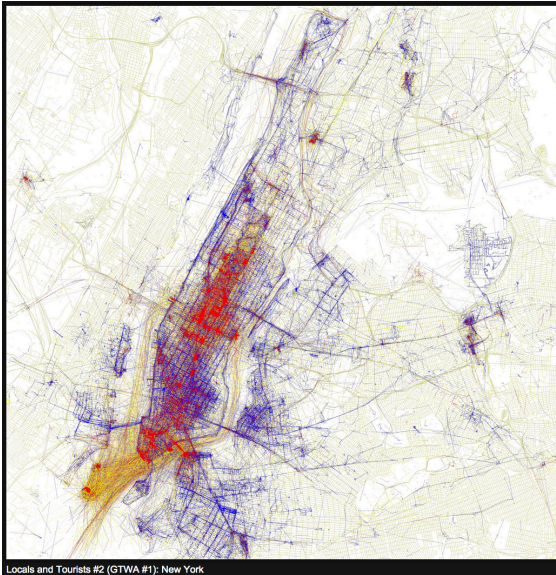


- 1 General info
- 2 Example: Baby names
- 3 Example: Geotagged data
- 4 Why study statistics?

### Clicker question

Do you geotag your posts on social networking sites, like Facebook, Twitter, Instagram, etc.?

- (a) yes
- (b) no



Locals and Tourists #2 (GTWA #1): New York  
<http://aaronstraupcope.com>

tourists

local

both



- 1 General info
- 2 Example: Baby names
- 3 Example: Geotagged data
- 4 Why study statistics?

## 'Lucky' woman who won lottery four times outed as Stanford University statistics PhD

By RACHEL QUIGLEY

Last updated at 7:02 PM on 9th August 2011

[Comments \(4\)](#) | [Add to My Stories](#) | [Share](#)

[Like](#) 2K

She was called the luckiest woman in the world.

But now that luck is being called into question by some who think that winning the lottery four times is more than just a coincidental spell of good fortune.

Joan R. Ginther, 63, from Texas, won multiple million dollar payouts each time.



Lady luck? Joan Ginther has won the Texas lottery four times, but is it remarkable good fortune or beating the system?



**Make a Difference**



**Have Fun**



**Satisfy Curiosity**



**Make Money**

---

**What do statisticians do? Here are a few examples:**

**Help animals**

Collaborate with other scientists to find ways to protect endangered species.

**Combat disease**

Help researchers understand how prevalent diseases are among various populations and why.

**Build winning professional sports teams**

Work for professional sports teams to help them pick the next season's new players.

**Make drugs safe**

Help develop new medicines that are safe and effective.

**Jobs in statistics are not only fun and exciting, but also smart for your future:**

**Job growth is strong**

Jobs for statisticians will grow **27 percent between 2012 and 2022**, much faster than the growth rate of 18 percent for computer occupations, and 11 percent for all occupations.

**Wages are high**

The median salary for data scientists with less than three years of experience is **\$80,000 and \$150,000** for those with nine or more years of experience, according to the Burtch Works 2014 report.