Chapter 1: Introduction

STAT 371

Duzhe Wang

the Department of Statistics, UW-Madison

Let's start off with...

"**Statistics** is a **Science** , not a branch of mathematics, but uses mathematical models as essential tools. "

-John Tukey

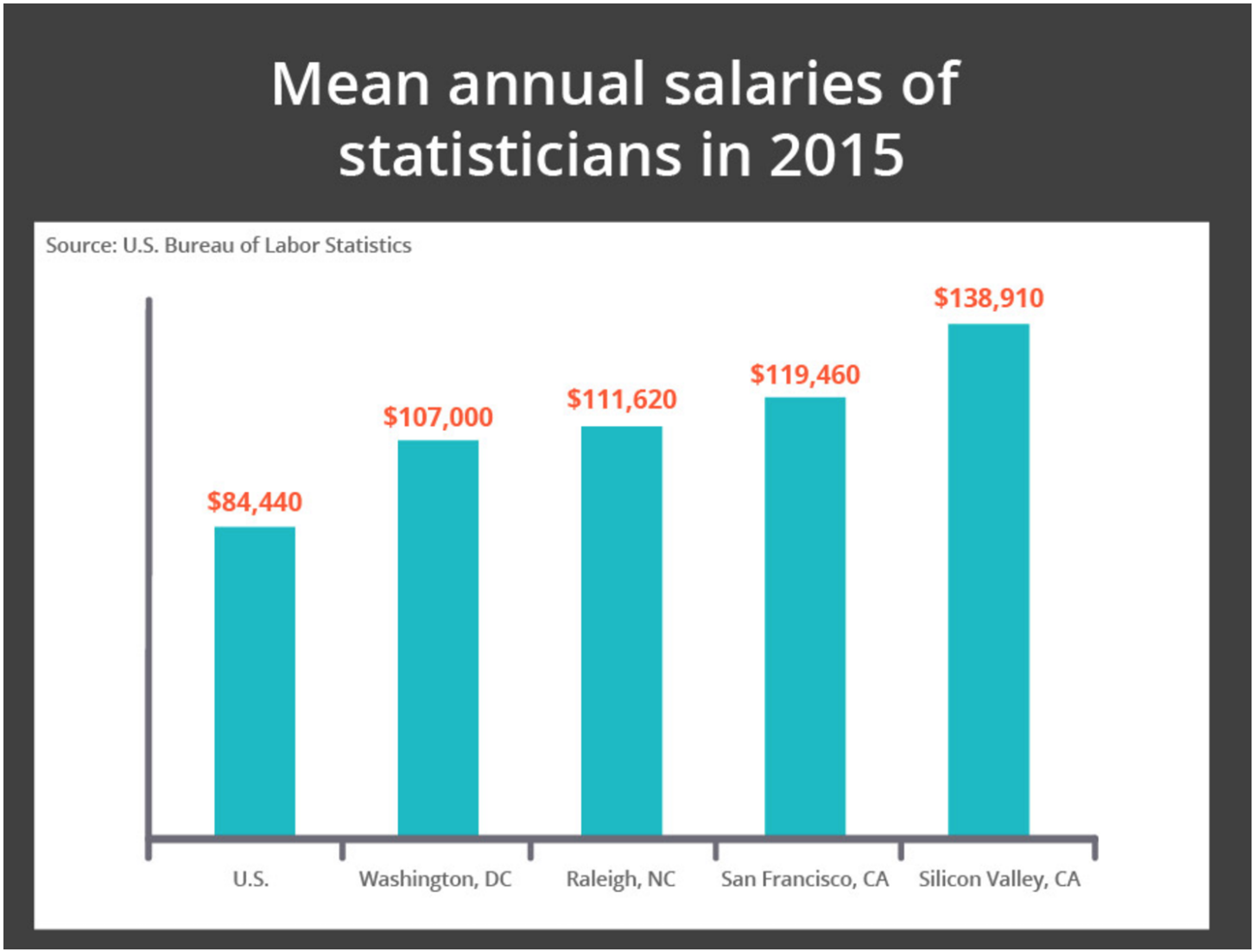
Outline

Why you need to study statistics

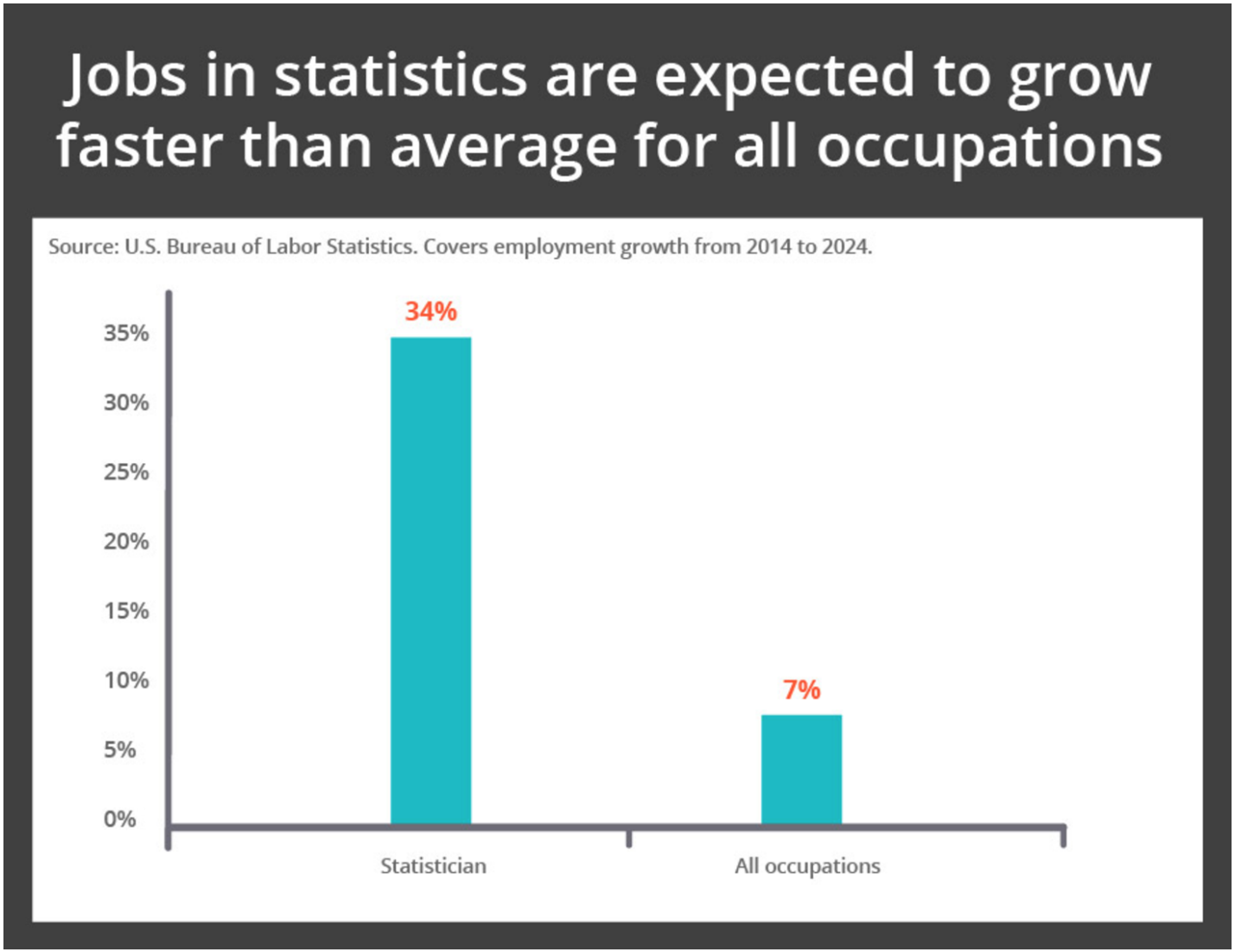
1. Make a difference.
   Statisticians contribute to society in many ways, from protecting endangered species and managing the impacts of climate change to making medicines more effective and reducing hunger and disease.
2. Have fun.
   After learning statistics, you could help professional sports teams pick the next season's new players, or a member of the data science team of a U.S. presidential campaign.
3. Make money.
   Demand for statisticians is growing, and so are their salaries. 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.

Outline

Let's look at some statistics



Let's look at some statistics, cont'd



Another interesting example

In 1973 the Graduate Division at The University of California, Berkeley conduced a study on sex bias in admissions.

Another interesting example

In 1973 the Graduate Division at The University of California, Berkeley conduced a study on sex bias in admissions.

	Male	Female
No. Applied	8442	4321
% Admitted	44%	35%

Another interesting example

In 1973 the Graduate Division at The University of California, Berkeley conduced a study on sex bias in admissions.

	Male	Female
No. Applied	8442	4321
% Admitted	44%	35%

"Assuming that the men and women were on the whole equally well qualified (and there is no evidence to the contrary), the difference in admission rates looks like a strong piece of evidence to show that men and women are treated differently in the admissions procedure."

Sex Bias in Graduate Admissions, cont'd

For graduate study, each major conducts its own admissions. What percentage of males vs. females are admitted by major?

Sex Bias in Graduate Admissions, cont'd

For graduate study, each major conducts its own admissions. What percentage of males vs. females are admitted by major?

Sex Bias in Graduate Admissions, cont'd

For graduate study, each major conducts its own admissions. What percentage of males vs. females are admitted by major?

	Men		Women	
Major	No. Applicants	% Admitted	No. Applicants	% Admitte
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

Sex Bias in Graduate Admissions, cont'd

For graduate study, each major conducts its own admissions. What percentage of males vs. females are admitted by major?

	Men		Women	
Major	No. Applicants	% Admitted	No. Applicants	% Admitte
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

Over 50% of men in this table applied to Majors A + B. Over 90% of women applied to Majors C-F.

Sex Bias in Graduate Admissions, cont'd

We see that sex of applicants is associated with the selectivity of different majors.

Sex Bias in Graduate Admissions, cont'd

We see that sex of applicants is associated with the selectivity of different majors.

When confronted with statistical data, it is often a good exercise to ask: are there any other explanations for these data that aren't being taken into consideration?

Outline

So what is Statistics about?

Statistics is about:

1. Summarizing data in "useful" ways, that could potentially reveal interesting patterns. This is a branch of Statistics known as **Descriptive Statistics.**

So what is Statistics about?

Statistics is about:

1. Summarizing data in "useful" ways, that could potentially reveal interesting patterns. This is a branch of Statistics known as **Descriptive Statistics.**
2. Determining if and to what extent patterns observed in data are "real," and generalize to a larger context. This is known as **Inferential Statistics.** In inferential statistics, the data form a **sample**, a smaller subset of some well-defined collection of things called a **population**. The idea is to use the sample to learn about the population–the process of inference.

What types of data do we encounter?

- Numerical: data that consist of numbers.

What types of data do we encounter?

- Numerical: data that consist of numbers.
  - Continuous: any value in a specified range is (theoretically) possible. Things we can measure to a near arbitrary degree of accuracy (with a sensitive enough instrument). E.g. time, weight, length.
  - Discrete: Only specific values are possible. E.g. number of children, any kind of count data.

What types of data do we encounter?

- Numerical: data that consist of numbers.
  - Continuous: any value in a specified range is (theoretically) possible. Things we can measure to a near arbitrary degree of accuracy (with a sensitive enough instrument). E.g. time, weight, length.
  - Discrete: Only specific values are possible. E.g. number of children, any kind of count data.
- Categorical: data that aren't numbers.

Course objectives

What you should get out of this course:

- Basic knowledge of probability and statistics, and why they are important.

Course objectives

What you should get out of this course:

- Basic knowledge of probability and statistics, and why they are important.
- The ability to summarize data graphically and numerically.

Course objectives

What you should get out of this course:

- Basic knowledge of probability and statistics, and why they are important.
- The ability to summarize data graphically and numerically.
- In some simple situations, the ability to identify and carry out the appropriate analysis.

Course objectives

What you should get out of this course:

- Basic knowledge of probability and statistics, and why they are important.
- The ability to summarize data graphically and numerically.
- In some simple situations, the ability to identify and carry out the appropriate analysis.
- In some simple situations, the ability to design an experiment that will yield the appropriate data.

Course objectives

What you should get out of this course:

- Basic knowledge of probability and statistics, and why they are important.
- The ability to summarize data graphically and numerically.
- In some simple situations, the ability to identify and carry out the appropriate analysis.
- In some simple situations, the ability to design an experiment that will yield the appropriate data.
- The ability to carry out simple analyses using a statistical computing package.

Course objectives

What you should get out of this course:

- Basic knowledge of probability and statistics, and why they are important.
- The ability to summarize data graphically and numerically.
- In some simple situations, the ability to identify and carry out the appropriate analysis.
- In some simple situations, the ability to design an experiment that will yield the appropriate data.
- The ability to carry out simple analyses using a statistical computing package.
- You won't leave fully fledged data analysts, but you will leave prepared to learn more about the discipline.