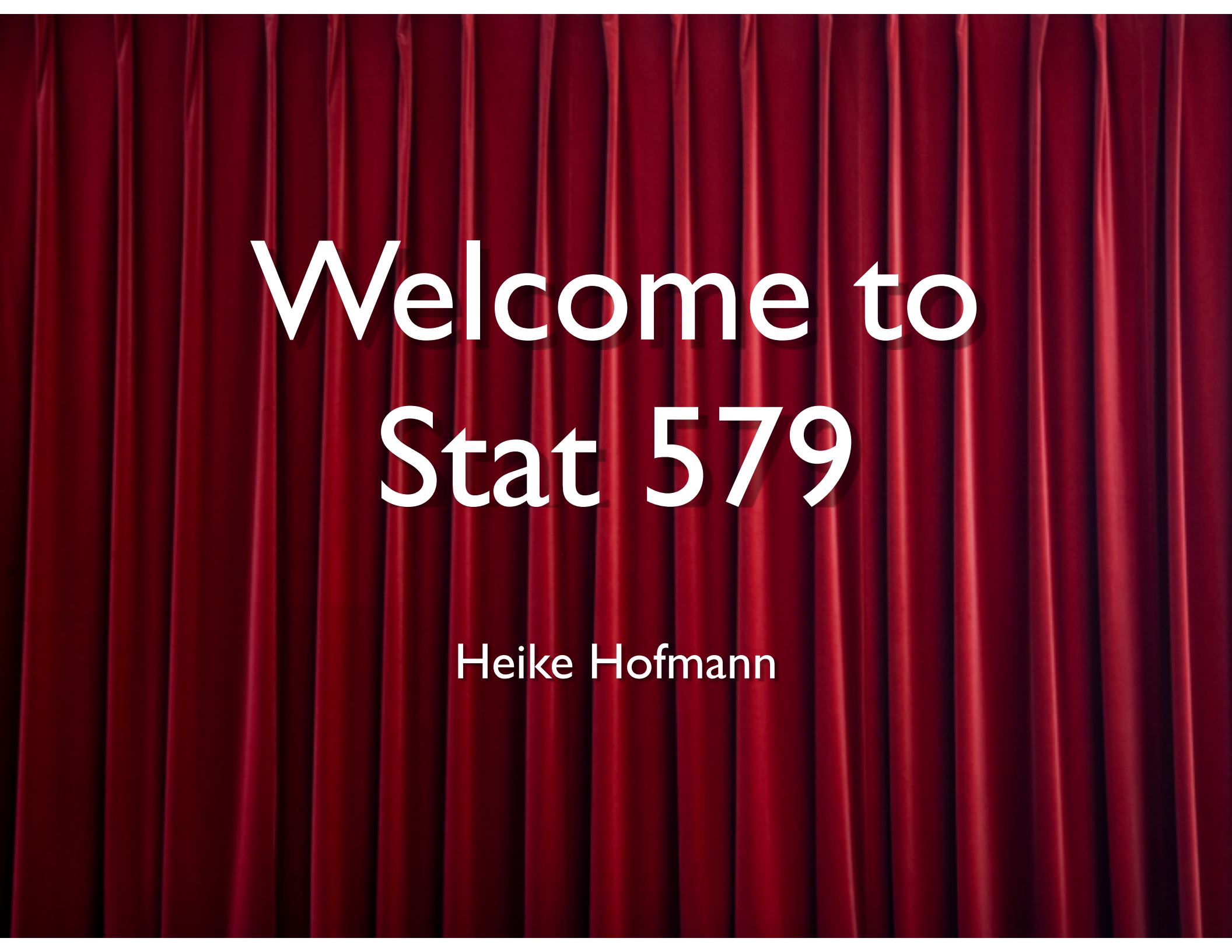
Welcome to
Stat 579
Heike Hofmann

About me ...

Professor of Statistics

my research area is exploratory data analysis and statistical computing, heavily involved in CSAFE

I have been teaching Stat 579 and other courses in statistical computing since 2006

About me ...

Professor of Statistics

my research area is exploratory data analysis and statistical computing, heavily involved in CSAFE

I have been teaching Stat 579 and other courses in statistical computing since 2006

About me ...

Professor of Statistics

my research area is exploratory data analysis and statistical computing, heavily involved in CSAFE

I have been teaching Stat 579 and other courses in statistical computing since 2006

easiest to reach at

hofmann@iastate.edu

easiest to reach at

hofmann@iastate.edu

R user since ~ 1995
R developer since ~ 1998

… what about you?

Before we get started ...

- course has two sections
  1 (Thursday, Statistics graduate students)
  2 (Tuesday, graduate students from outside)

- time: 12:10 - 2:00

You all have different backgrounds ...

- Ideally, Stat 579 helps you learn *more* R

- Tuesday section: introductory R
- Thursday section: intermediate R

You choose
Mix & match is OK

- I will record the class and post videos in Canvas

Where do I find …

- Class slides, assignments, code
  https://stat579-at-isu.github.io/
- Lecture videos, grades
  Canvas

Outline

- Introduction
- Objectives, software, modules
- Assessment/Grades
- What do you already know?
- What are you interested in?
- Getting started

“Tall-order” Objective

- become computationally proficient to do statistical data analysis

Objectives

- Learn how to deal with complex, messy, real data
- Use graphics to explore and understand data
- Gain familiarity with basic data collection, storage and manipulation
- Fluently reshape data into the most convenient form for analysis or reporting
- Automate cleaning and analysis

Additional Readings

- "R For Data Science" by Hadley Wickham and Garret Grolemund at https://r4ds.hadley.nz/

- Recommended: "ggplot2" by Hadley Wickham (email me for the pdf or use Springer link for *use R!* series)

Setup of Slides

Learn about **Concepts** of Statistical Computing
Practice **Skills**

about 2:1 split of time in reading vs practice,
set an alarm to see whether you can solve your turns in that amount of time

Software: R

Assessments

- One homework assignment every week + quizzes to test understanding of basics
- One midterm exam
- final (team) project & presentation

- 30% homework, 5% quizzes/surveys,
  30% midterm, 35% final project
  (25% writeup, 10% presentation)
- extra credit is listed as 0% … so it is extra

Homework

- One every week
- 2-3 hours
- Revise what we covered, plus synthesize some new information

- I will publish the best answers on the class website anonymously, if that's not ok, please note that on your assignment

Midterm

- during our regular meeting time …
- Open note, open book, open internet … but no collaboration with anybody

- tentatively scheduled for the last week of October

Final Project

- 1 bigger project (multiple due dates)
- Will be open ended, and will involve a substantial write up (~10 pages)
- Work in a team (3-4 members)
- Talk to each other! Find a group!

Disability/Sickness

- Make sure to let me know (in advance) if you need any accommodations

- If you are sick, please take care of yourself
- Catch up - make use the notes, your friends, and of office hours.

Lectures

- Electronic copy of the slides will be available on the website
- But you'll need to take your own notes!
  - If you really want complete notes, organize a roster with others in the class
- Don't goof off on the computers!
- If you're bored, complain!

What do you know already?

- Have you used R before?
- … for more than a year?
- the tidyverse?
- RMarkdown?
- git?
- … there will be a survey on Canvas …

What is this class about?

- Very Data Centric
- I want to know about topics you're interested in

- Sports
  e.g. Baseball salaries and performance
- Crime data (incl. type, time, place, demographics etc.)
- Health
  e.g. fitness statistics, or disease rates, or health care
- Movies
  e.g. ratings/box office revenues from IMDB
- Climate/Weather Data
- Travel data, e.g. US flights
- National or Global Issues (world poverty, carbon footprint, …)

- Anything else you can think of?

Go to Wufoo Site to let me know your favorites and make suggestions:
https://heike.wufoo.com/forms/topics-of-interest/