

# ISOM 3390: Business Programming in R

## Topic 1: Course Introduction

Bingjie QIAN (PhD candidate)

Summer 2020, HKUST

# What is R?

- R is a programming language that was developed in 1993 by Robert Gentleman and Ross Ihaka in the University of Auckland, New Zealand.
- A powerful tool to retrieve, clean, analyze and visualize data.
- One of the most popular programming languages among statisticians, data scientists, marketers and researchers.
- Widely used in statistical computing, business analytics and scientific research.

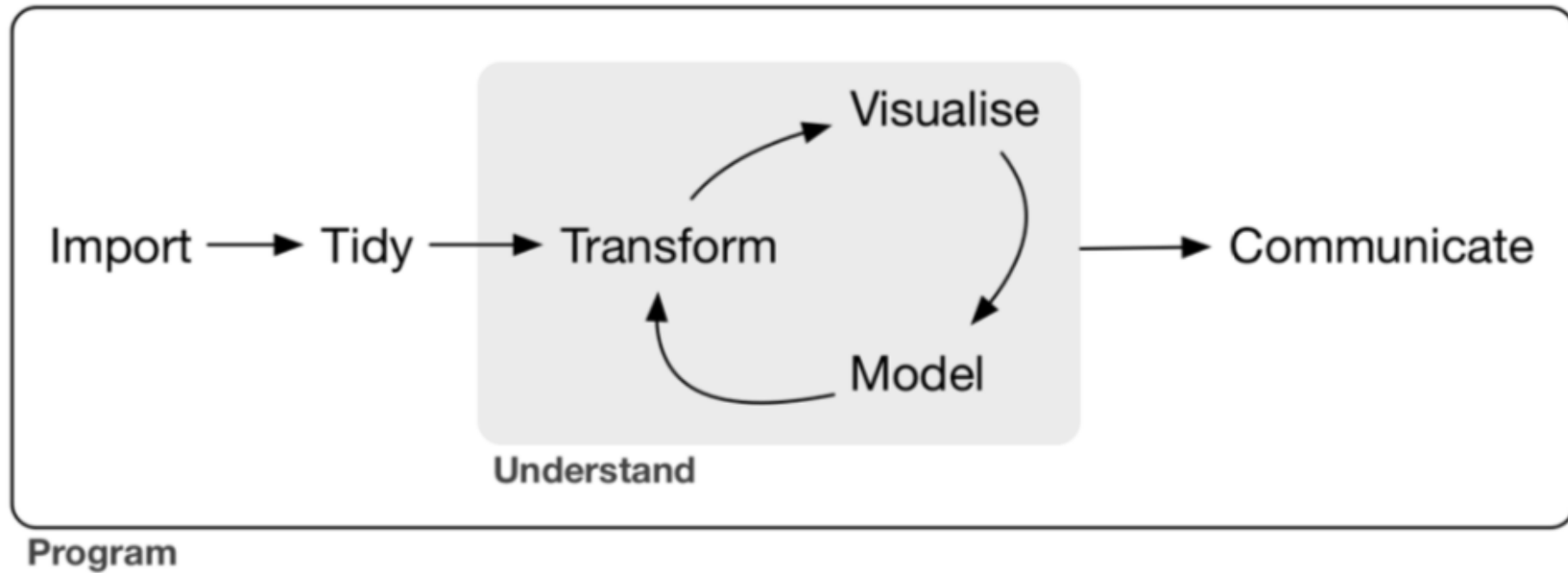
# Why Use R?

- R is open-source and free
- R is popular, and increasing in popularity
- R runs on all platforms: Windows, Mac, Linux
- R is powerful and highly extensible (via packages)

# Is R Difficult to Learn?

- On the one hand,
  - Programmers that come from a Python, PHP or Java background might find R quirky and confusing at first.
  - The syntax that R uses is a bit different from other common programming languages.
  - There are too many ways to do similar things: different functions from different packages.
- On the other hand,
  - R's syntax is *expressive*, allowing users – even those from non computer science backgrounds to quickly import, clean and analyze data from various data sources.
  - R has a rich collection of packages, providing the broadest range of statistical and data analysis tools, so you don't have to do everything from scratch.

# What You will Learn



# What You will Learn

- Scraping data from the web: Trump's tweets



**Donald J. Trump**

@realDonaldTrump

I want to do negative ads on John Kasich, but he is so irrelevant to the race that I don't want to waste my money.

5,597 9:05 AM - Nov 20, 2015

2,945 people are talking about this

# What You will Learn

- Tidying and manipulating data

Tweet ID	Text	Date	Favorites	
759592590106849280	Nielson Media Research final numbers on ACCEPTANCE SPEECH: TRUMP 32.2 MILLION. CLINTON 27.8 MILLION. Thank you!	2016-07-30T23:32:40Z	13850	4130
759524001613918208	Thank you to all of the television viewers that made my speech at the Republican National Convention #1 over Crooked Hillary and DEMS.	2016-07-30T19:00:07Z	27659	6842
759516008272932864	Can you imagine if I had the small crowds that Hillary is drawing today in Pennsylvania. It would be a major media event! @CNN @FoxNews	2016-07-30T18:28:22Z	19968	6488
759515080010719232	NATO commander agrees members should pay up via @dcexaminer:http://www.washingtonexaminer.com/nato-commander-agrees-members-should-pay-up/article/2598183?custom_click=rss ...	2016-07-30T18:24:40Z	11624	4668
759513644258525184	Wow, NATO's top commander just announced that he agrees with me that alliance members must PAY THEIR BILLS. This is a general I will like!	2016-07-30T18:18:58Z	23922	7819
759512042906587137	The "Rust Belt" was created by politicians like the Clintons who allowed our jobs to be stolen from us by other countries like Mexico. END!	2016-07-30T18:12:36Z	17239	5717
759509936002957314	Crooked's stop in Johnstown, Pennsylvania, where jobs have been absolutely decimated by dumb politicians, drew less than 200 - with Bill, VP	2016-07-30T18:04:14Z	19254	7013
759507812963737600	#CrookedHillary = Obama's third term, which would be terrible news for our economic growth - seen below.pic.twitter.com/y9WJoUaaql	2016-07-30T17:55:48Z	15358	6439

Showing 1 to 8 of 4,654 entries

Previous

1

2

3

4

5

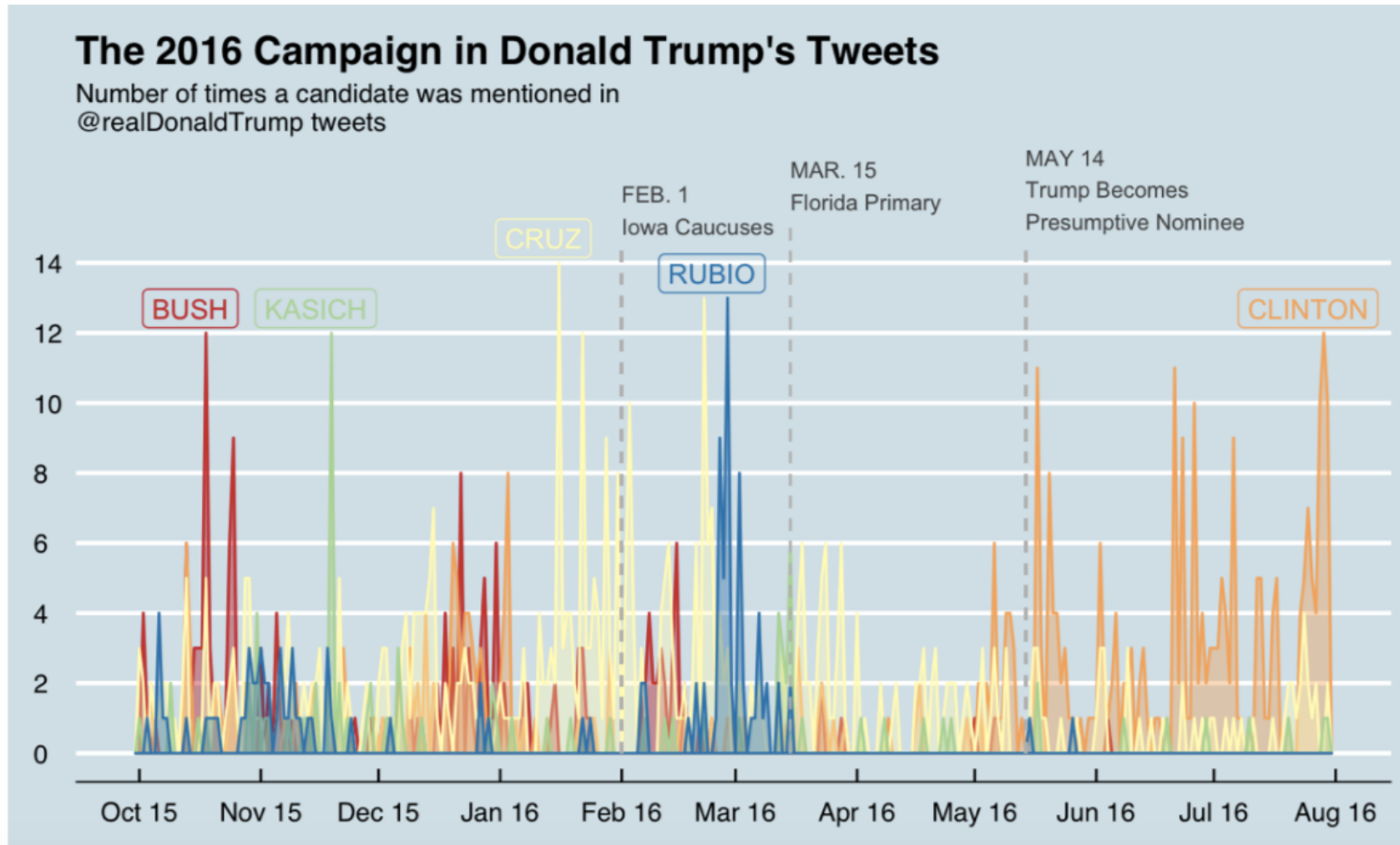
...

582

Next

# What You will Learn

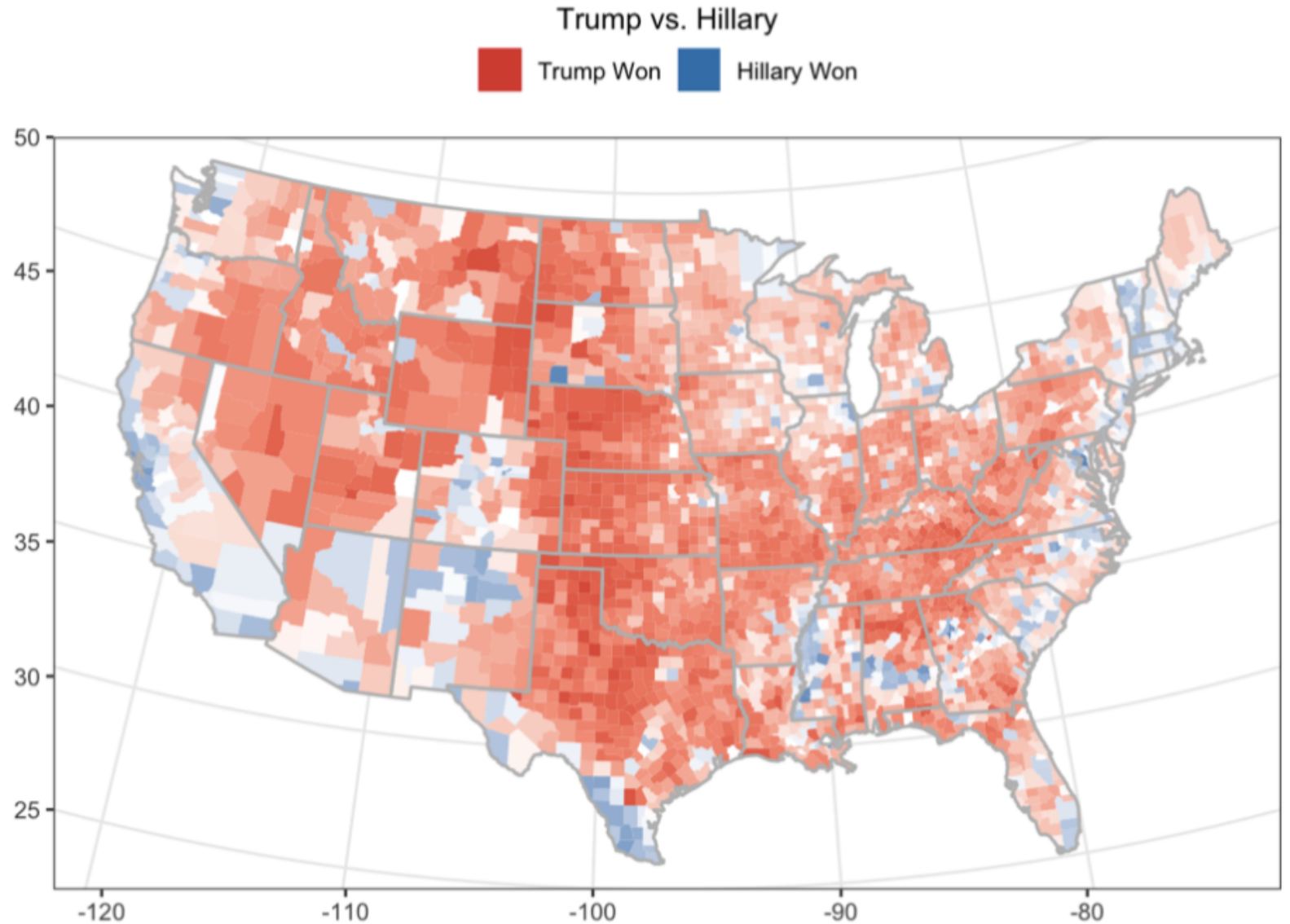
- Visualizing data





# What You will Learn

- Visualizing data
  - 2016 US presidential election



# What You will Learn

- Text analytics: understand subtle and nuanced emotions



## **The Favourite** (2018)

★ 10/10

**I was left shaking after watching this film**

5 November 2018

Honestly all I can say is that this film was not what I was expecting and far exceeded my expectations. The chemistry between the actors and also the visual story is absolutely stunning and I'm just wowed by how well done everything is done in this film. I can't say I have anything bad to say about this film. And please go into this movie without spoilers, I find that it is way more enjoyable to be surprised by the actual story and leaves more excitement for the viewer.



## **La La Land** (2016)

★ 4/10

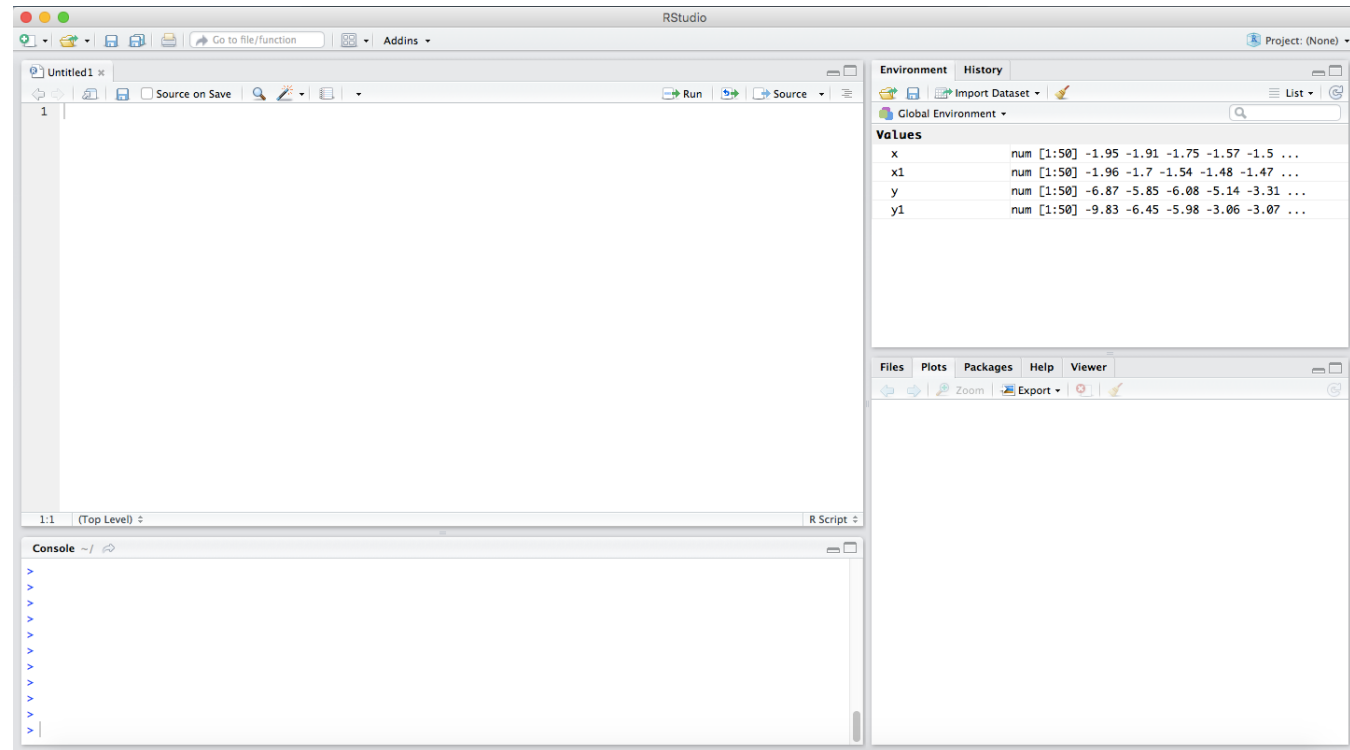
**average, but keep trying Hollywood**

16 January 2017

The dancing was average at best. The opening scene was the best routine. These actors are not Ginger and Fred. I am not sure if the talent was missing, or something else. The singing was either weak, or mixed poorly. The song writing was unique, but not strong enough to remember. The story line was predictable but sweet. I was so looking forward to seeing a musical movie. It has received great reviews because I believe those reviewers were all so desperate to see an upbeat, musical love story, as was I. So, in conclusion, it is worth seeing, if only to just to send Hollywood a message, with your dollars. More of this but better quality please.

# RStudio

- RStudio is an integrated development environment (IDE) for R.
- *A tour on RStudio*



# R Markdown



- An R Markdown file is a plain-text file that has the extension **.Rmd**
- R Markdown provides a productive notebook interface and a unified authoring framework that weave together **plain text and code** to produce an elegantly formatted output.
- R Markdown support dozens of output formats, like PDFs, Word files, slideshows, and more.
- *An example:* RMarkdown\_Demo.Rmd (download from Canvas)

# R Markdown

- An R Markdown file contains three important types of content:
  - An (optional) YAML header surrounded by `---`.
  - Chunks of R code surrounded by `````.
  - Text mixed with simple text formatting like `# heading` and `*italics*`.
- For more information, in RStudio:
  - Help → Markdown Quick Reference
  - Help → Cheatsheets → R Markdown Cheat Sheet
- Click "Knit" to produce a complete report containing all text, code, and results.

# Course Information

- Lecture
  - Time: Mon & Wed & Fri 1:30-4:20pm
  - Zoom meeting ID: 996-5043-1367
  - Instructor: Bingjie Qian (PhD candidate)
  - Email: [bqianaa@connect.ust.hk](mailto:bqianaa@connect.ust.hk)
  - Office: LSK 4063
- Lab
  - Time: Mon & Wed & Fri 4:20-5:20pm
  - Zoom meeting ID: 949-5602-2516
  - Instructor: Josh
  - Email: [imjosh@ust.hk](mailto:imjosh@ust.hk)
  - Office: LSK 4049C

# Course Information

- Grading

1. In-class exercises (13): 20%
2. Lab exercises (13): 20%
3. Assignments (2): 30%
  - Individual homework
  - Longer and more complex tasks
4. Final Project (group): 30%
  - Group size: at maximum 3 per group;
  - 10% out of 30% will be based on your teammates' assessment.

# Class Policy

- Feel free to interrupt me and ask questions if you feel confused.
- Get your hands dirty, you need a lot of practice!!
- Participating in live meetings is highly recommended, but you can also watch the recorded videos to learn.
  - Recorded videos are available upon request.
  - Submit in-class and lab exercises before due (11:59am the next day).
- For each topic, I will create a Discussion on Canvas, where you can post your questions. In-person consultation will be arranged on case-by-case basis.
- To contact me, please email me directly instead of sending messages through Canvas Inbox.



Questions?