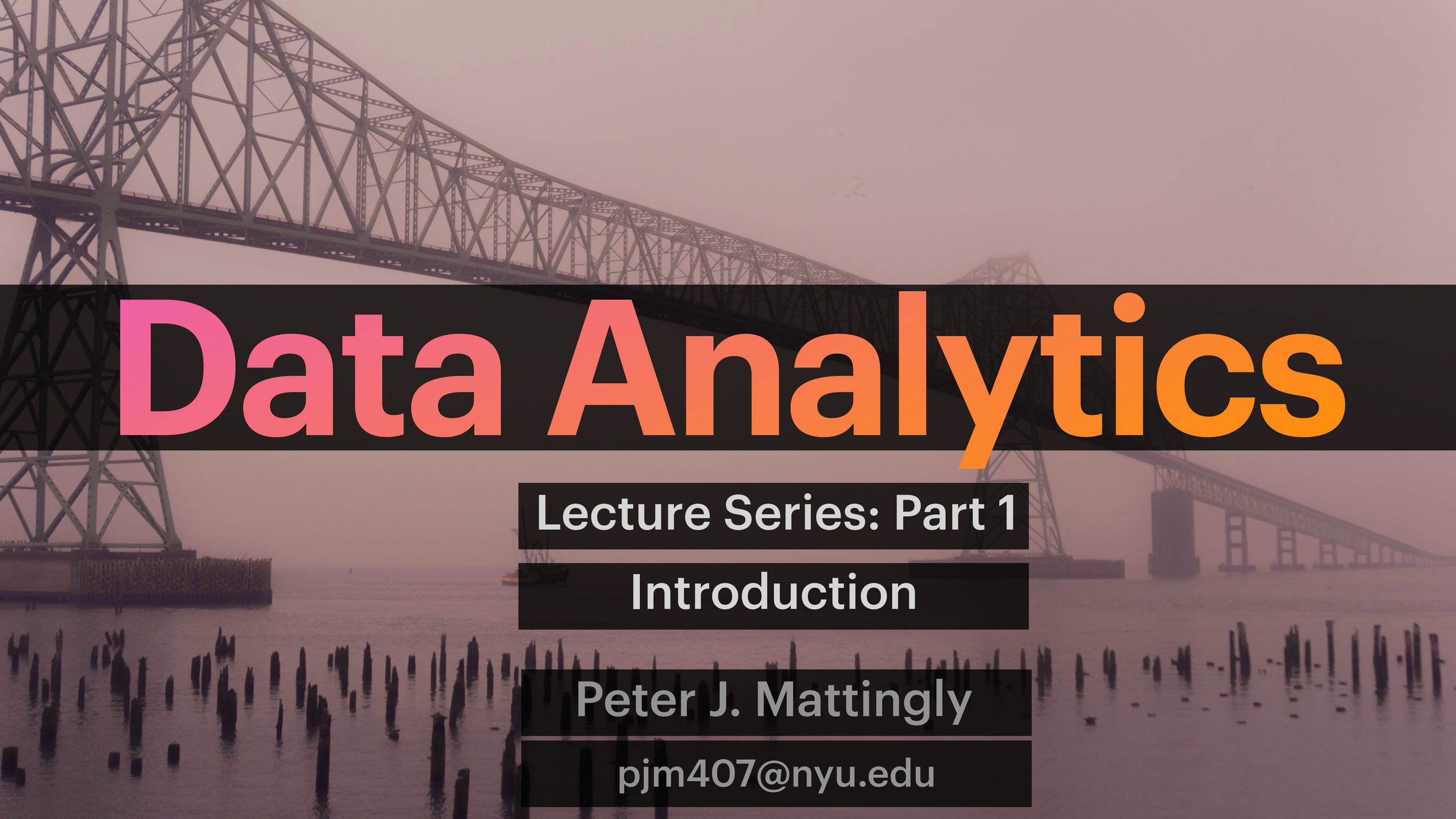


Data Analytics

Lecture Series: Part 1

Introduction

Peter J. Mattingly

pjm407@nyu.edu

Welcome!

Welcome!

In the course, we will:

Welcome!

In the course, we will:

- Source and analyze cities data for real estate analysis

Welcome!

In the course, we will:

- Source and analyze cities data for real estate analysis
- Create reproducible working environments

Welcome!

In the course, we will:

- Source and analyze cities data for real estate analysis
- Create reproducible working environments
- Visualize insight and information

Overview

In this section, we will:

Overview

In this section, we will:

- Define data analytics

Overview

In this section, we will:

- Define data analytics
- Introduce R Studio


@tladeras

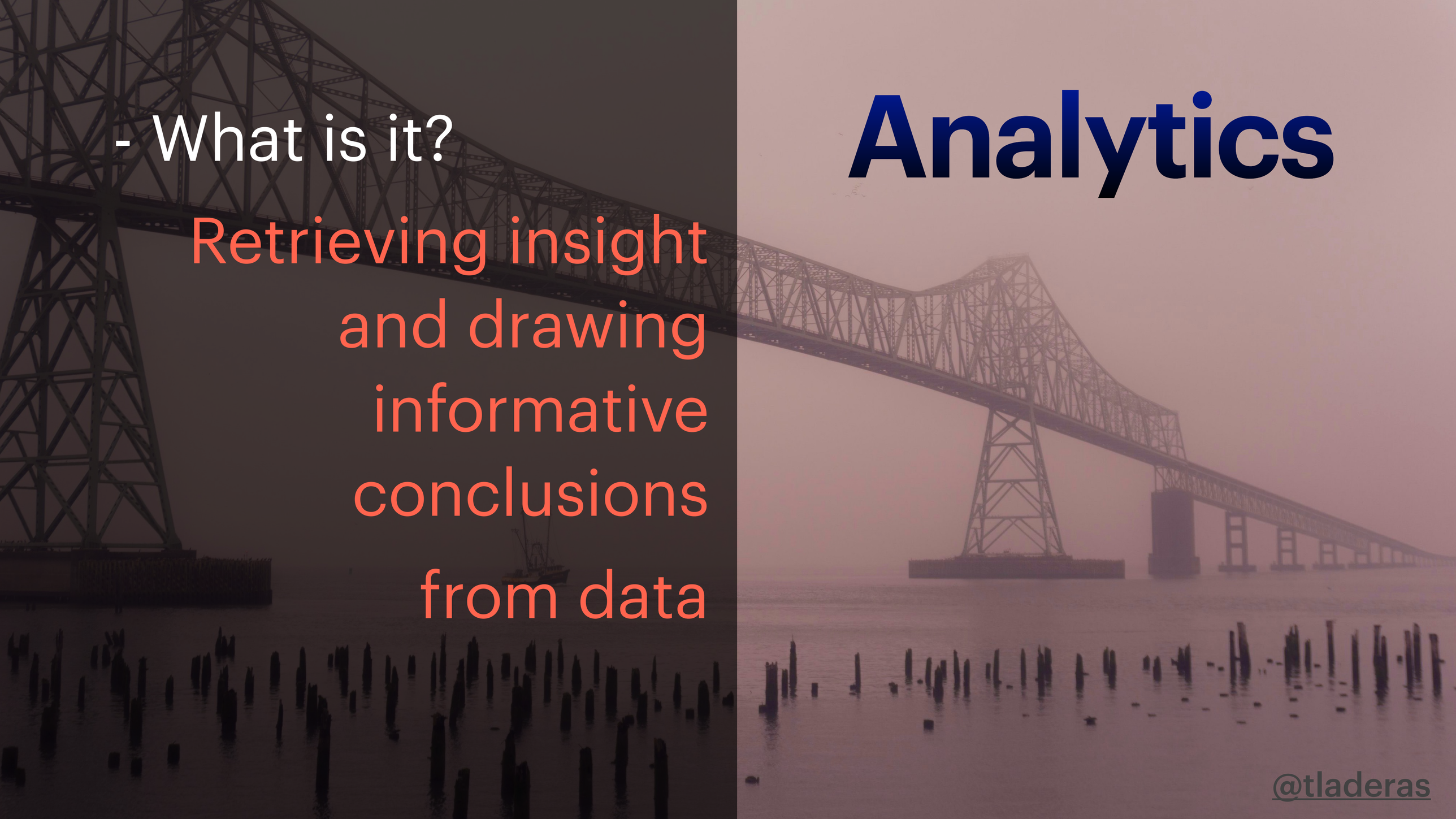
Analytics

@tladeras

Analytics

@tladeras


- What is it?
Analytics
@tladeras



Analytics

- What is it?

Retrieving insight and drawing informative conclusions from data

@tladeras


- What data?
Analytics
@tladeras


- What data?
Imported, cleaned, transformed, visualized, and modeled data
Analytics
@tladeras


Data
@tladeras

Data

	GeoFips	GeoName	2007	2010	2015
1	998	United States (Metropolitan Portion)	12821473605	13265807959	16264292775
2	10180	Abilene, TX (Metropolitan Statistical Area)	5406560	5861966	6836185
3	10420	Akron, OH (Metropolitan Statistical Area)	28749370	28922844	34122942
4	10500	Albany, GA (Metropolitan Statistical Area)	5101790	5109457	5434079
5	10540	Albany-Lebanon, OR (Metropolitan Statistical Area)	3643627	3460440	4116193
6	10580	Albany-Schenectady-Troy, NY (Metropolitan Statistical…	42731610	47044326	54546375
7	10740	Albuquerque, NM (Metropolitan Statistical Area)	35108392	36388647	39419488
8	10780	Alexandria, LA (Metropolitan Statistical Area)	4636424	5369983	5862965
9	10900	Allentown-Bethlehem-Easton, PA-NJ (Metropolitan Sta…	33647001	35746169	41542046
10	11020	Altoona, PA (Metropolitan Statistical Area)	4643209	4962165	5547413
11	11100	Amarillo, TX (Metropolitan Statistical Area)	9788225	10722023	12784064

@tladeras

Data

	GeoFips	GeoName	2007	2010	2015
1	998	United States (Metropolitan Portion)	12821473605	13265807959	16264292775
2	10180	Abilene, TX (Metropolitan Statistical Area)	5406560	5861966	6836185
3	10420	Akron, OH (Metropolitan Statistical Area)	28749370	28922844	34122942
4	10500	Albany, GA (Metropolitan Statistical Area)	5101790	5109457	5434079
5	10540	Albany-Lebanon, OR (Metropolitan Statistical Area)	3643627	3460440	4116193
6	10580	Albany-Schenectady-Troy, NY (Metropolitan Statistical…	42731610	47044326	54546375
7	10740	Albuquerque, NM (Metropolitan Statistical Area)	35108392	36388647	39419488
8	10780	Alexandria, LA (Metropolitan Statistical Area)	4636424	5369983	5862965
9	10900	Allentown-Bethlehem-Easton, PA-NJ (Metropolitan Sta…	33647001	35746169	41542046
10	11020	Altoona, PA (Metropolitan Statistical Area)	4643209	4962165	5547413
11	11100	Amarillo, TX (Metropolitan Statistical Area)	9788225	10722023	12784064

@tladeras

Data

	GeoFips	GeoName	2007	2010	2015
1	998	United States (Metropolitan Portion)	12821473605	13265807959	16264292775
2	10180	Abilene, TX (Metropolitan Statistical Area)	5406560	5861966	6836185
3	10420	Akron, OH (Metropolitan Statistical Area)	28749370	28922844	34122942
4	10500	Albany, GA (Metropolitan Statistical Area)	5101790	5109457	5434079
5	10540	Albany-Lebanon, OR (Metropolitan Statistical Area)	3643627	3460440	4116193
6	10580	Albany-Schenectady-Troy, NY (Metropolitan Statistical…	42731610	47044326	54546375
7	10740	Albuquerque, NM (Metropolitan Statistical Area)	35108392	36388647	39419488
8	10780	Alexandria, LA (Metropolitan Statistical Area)	4636424	5369983	5862965
9	10900	Allentown-Bethlehem-Easton, PA-NJ (Metropolitan Sta…	33647001	35746169	41542046
10	11020	Altoona, PA (Metropolitan Statistical Area)	4643209	4962165	5547413
11	11100	Amarillo, TX (Metropolitan Statistical Area)	9788225	10722023	12784064

@tladeras

Data

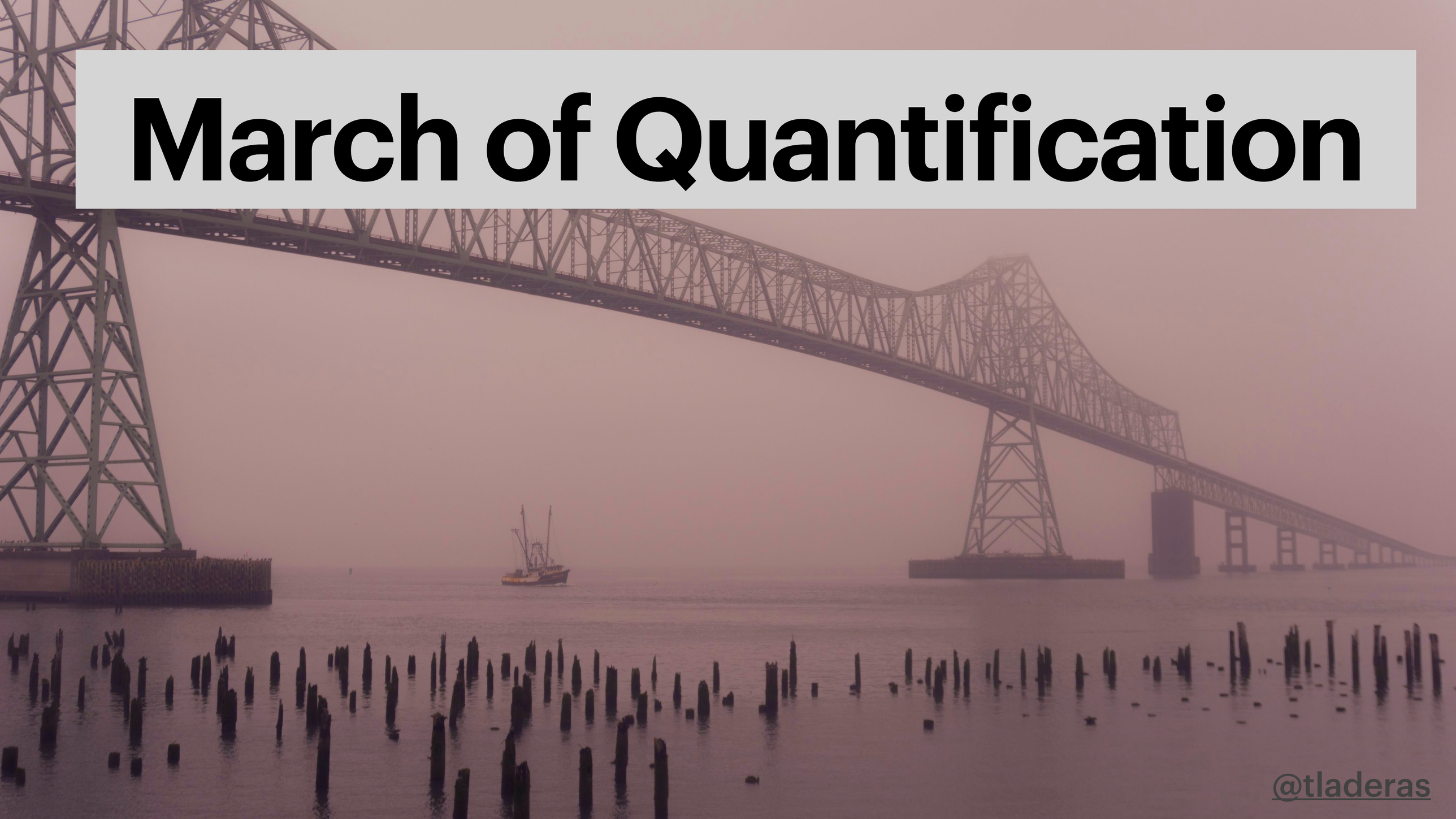
	GeoFips	GeoName	2007	2010	2015
1	998	United State (Metropolitan Portion)	12821473605	13265807959	16264292775
2	10180	Abilene, TX (Metropolitan Statistical Area)	5406560	5861966	6836185
3	10420	Akron, OH (Metropolitan Statistical Area)	28749370	28922844	34122942
4	10500	Albany, GA (Metropolitan Statistical Area)	5101790	5109457	5434079
5	10540	Albany-Lebanon, OR (Metropolitan Statistical Area)	3643627	3460440	4116193
6	10580	Albany-Schenectady-Troy, NY (Metropolitan Statistical…	42731610	47044326	54546375
7	10740	Albuquerque, NM (Metropolitan Statistical Area)	35108392	36388647	39419488
8	10780	Alexandria, LA (Metropolitan Statistical Area)	4636424	5369983	5862965
9	10900	Allentown-Bethlehem-Easton, PA-NJ (Metropolitan Sta…	33647001	35746169	41542046
10	11020	Altoona, PA (Metropolitan Statistical Area)	4643209	4962165	5547413
11	11100	Amarillo, TX (Metropolitan Statistical Area)	9788225	10722023	12784064

@tladeras


- Why now?
Analytics
@tladeras


- Why now?
march of quantification
Analytics
@tladeras

March of Quantification



@tladeras

March of Quantification

Across all fields, disciplines, and industries :

March of Quantification

Across all fields, disciplines, and industries :

- Data more accessible via Internet & APIs

March of Quantification

Across all fields, disciplines, and industries :

- Data more accessible via Internet & APIs
- Methodologies more sophisticated & idiosyncratic

March of Quantification

Across all fields, disciplines, and industries :

- Data more accessible via Internet & APIs
- Methodologies more sophisticated & idiosyncratic
- Software more advanced (e.g., R, Python, Stata)

March of Quantification

Across all fields, disciplines, and industries :

- Data more accessible via Internet & APIs
- Methodologies more sophisticated & idiosyncratic
- Software more advanced (e.g., R, Python, Stata)
- Still need to uncover more


@tladeras


- Analyze data with what?
Analytics
@tladeras


- Analyze data with what?
software
Analytics
@tladeras


Software
@tladeras


Excel
Software
@tladeras

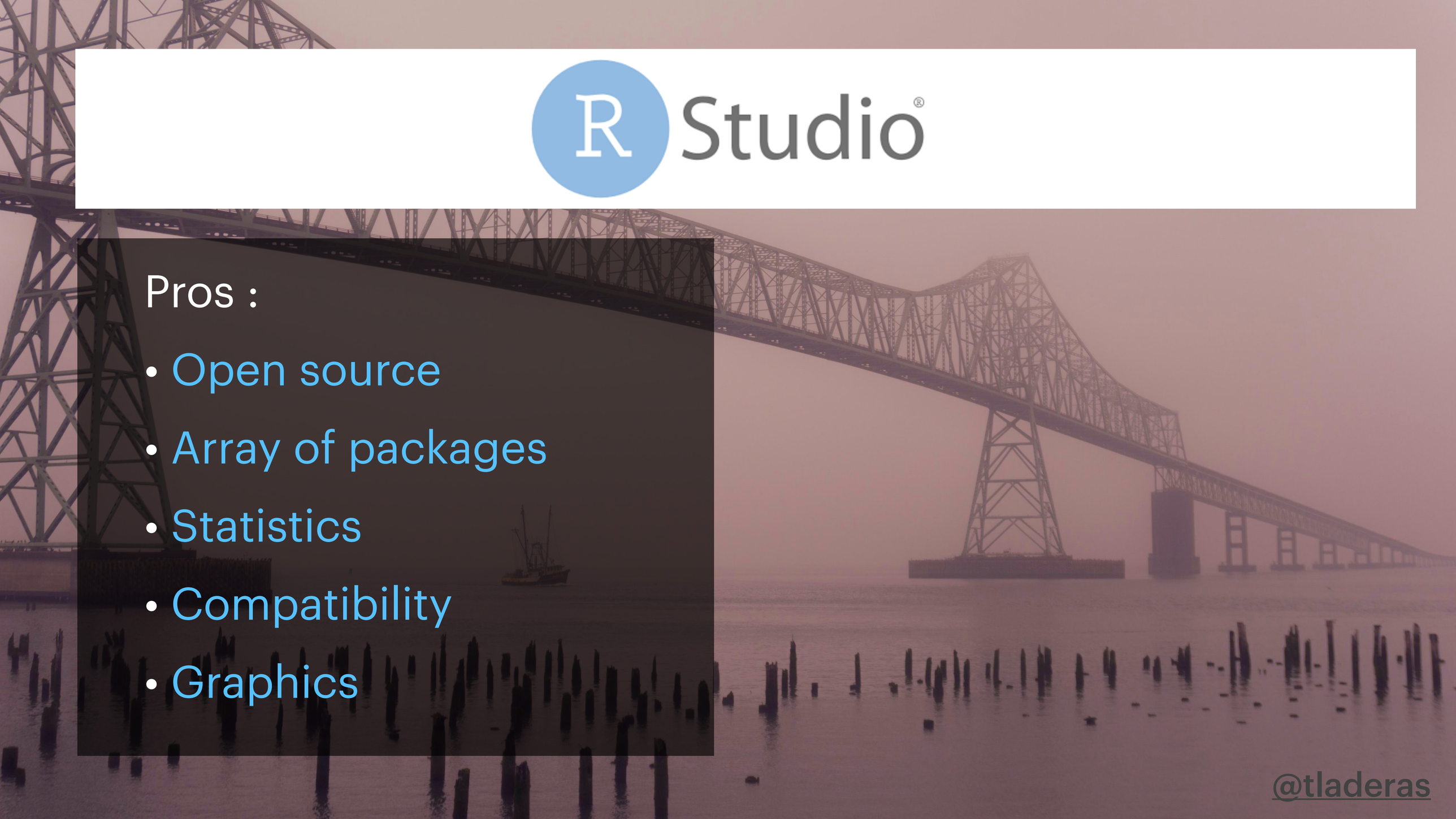

Excel
STATA®
Software
@tladeras


Excel
STATA®
python
Software
@tladeras


- What’s the best?
Analytics
@tladeras


RStudio®
@tladeras


RStudio®
Pros :
@tladeras



Pros :

- Open source
- Array of packages
- Statistics
- Compatibility
- Graphics

@tladeras

Pros :

- Open source
- Array of packages
- Statistics
- Compatibility
- Graphics

Cons :

@tladeras

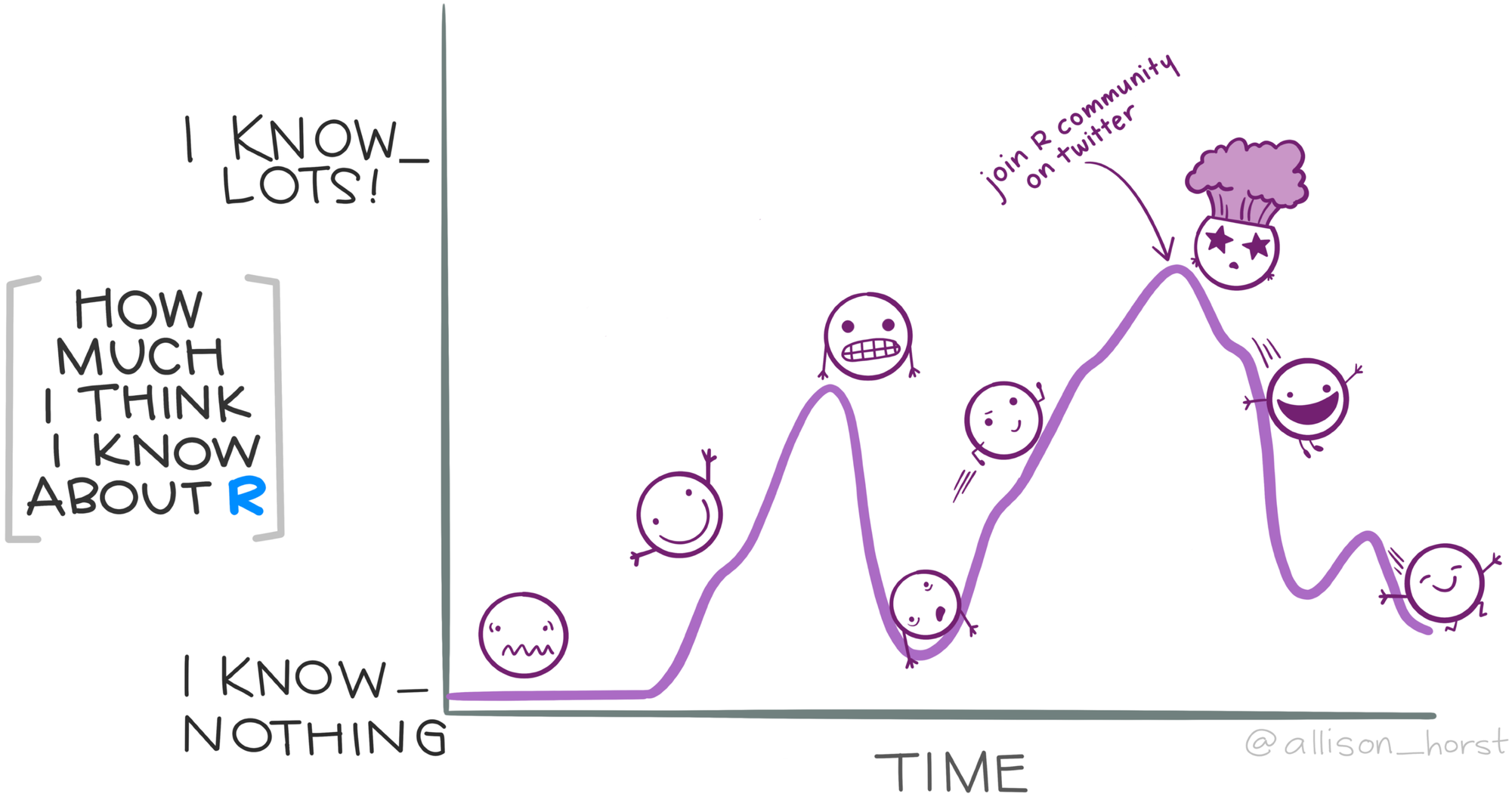
Pros :

- Open source
- Array of packages
- Statistics
- Compatibility
- Graphics

Cons :

- Memory
- Array of packages
- Security
- Learning curve

@tladeras


I KNOW_ LOTS!
HOW MUCH I THINK I KNOW ABOUT R
I KNOW_ NOTHING
join R community on twitter
TIME
@allison_horst


@tladeras