

# Welcome to the IUSSP Workshop on Social Media and Demographic Research

**Emilio Zagheni & Connor Gilroy**

University of Washington, Seattle

International Population Conference 2017, Cape Town

Many Thanks to IUSSP and Hewlett Foundation  
for their support

# IUSSP Panel on Big Data and Population Processes



Bogdan State



Emilio Zagheni



Francesco Billari



Emmanuel Letouze



Guy Abel



Melinda Mills



Cassio Turra



Mary Ellen Zuppan (IUSSP)



Ingmar Weber



Tom Le Grand



Paul Monet  
(IUSSP Secretariat)

# Recent activities of the IUSSP Panel

## ► **Hands-on Training Workshops**

- at Population Association of America (2016 and 2017)
- at European Population Conference (2016)
- at Brazilian+Latin American Population Conference (2016)

## ► **Mini-conferences/Research Workshops**

- at the International Conference on Web and Social Media (2016 and 2017)
- at Social Informatics Conference (2016)

## ► **Tutorials for (mainly) computer scientists**

- Tutorial on ‘Digital Demography’ at the World Wide Web Conference (2017) and at the International Conference on Computational Social Science (2017)

# Lots of people contributed beyond the IUSSP panel...

- ▶ Directly by preparing and/or teaching modules :
  - Monica Alexander, Nina Cesare, Lee Fiorio, Charles Lanfear, Kivan Polimis, Joan Ryan, ...
- ▶ Indirectly by developing packages for R and Python
  - E.g., Pablo Barbera, Matheus Araujo

## Links to available materials

1. European Population Conference 2016 workshop:

<https://github.com/CSDE-UW/iussp-mainz-social-media>

2. Latin American Pop Conference 2016 workshop:

<https://github.com/ezagheni/IUSSP-Social-Media-Workshop-at-ALAP>

3. Tutorial at WWW Conference 2017:

<https://t.co/4HJjeVgIEr>

# Upcoming Research Workshop



The workshop will be held in July in Shanghai, as a side meeting of APA

# General goals for today's workshop

3 broad questions:

1. How to collect and analyze social media data?
2. How can social scientists benefit from and contribute to social media research?
3. What are some examples of research questions in the emerging area of 'digital demography'?

⇒ Emphasis on having fun with hands-on activities

Not everything runs smoothly the first time...





...but eventually we will get there



# Plan for today

1. Introduction to APIs (Application Programming interfaces)
2. Collecting Twitter data using R
3. Facebook Data for Advertisers: a new source and research examples
4. Collecting Facebook Data for Advertisers
5. Discussion

# Materials for the workshop

- ▶ Materials for today's workshop are available on this github page:

`https://github.com/CSDE-UW/  
IUSSP-digital-demog-2017`

# Four main modules

8:30-9:00	Introduction
9:00-10:00	Lab: APIs
10:00-10:15	Break
10:15-11:45	Lab: Twitter
11:45-12:45	Lunch
12:45-13:30	Lab: Twitter + Visualization
13:30-14:15	Using Facebook Ads for research
14:15-14:30	Break
14:30-15:30	Lab: Facebook Ads
15:30-16:00	Discussion

## Modules

The linked folders and repositories contain the code and slides for the hands-on modules.

Module 1	<a href="#">Retrieving data through APIs</a>
Module 2	<a href="#">Collecting and analyzing Twitter data</a>
Module 3	<a href="#">Visualizing and mapping Twitter data</a>
Module 4	<a href="#">Demographic estimates from the Facebook Marketing API</a>

1: University of Washington

# Download the material and unzip the folder for each module

ccgilroy / retrieving-data-through-apis

Watch 0

Star 0

Fork 1

Code

Issues 0

Pull requests 0

Projects 0

Wiki

Insights

Several examples of accessing data through APIs using httr in R, with a focus on demographic data

7 commits

1 branch

0 releases

1 contributor

Branch: master

New pull request

Create new file

Upload files

Find file

Clone or download

ccgilroy update census endpoint, fix typo

Latest commit 710adb6 3 days ago

.gitignore	initial commit	14 days ago
EXAMPLE_census_api_key.txt	initial commit	14 days ago
README.md	update census endpoint, fix typo	3 days ago
working_with_apis.R	update census endpoint, fix typo	3 days ago
working_with_apis.Rproj	initial commit	14 days ago
working_with_apis.pdf	update census endpoint, fix typo	3 days ago

README.md

## Retrieving data through APIs

Drinking from a hose...



# Intro to accessing data via APIs

Among others, we will consider data like this...



TWEETS 17 FOLLOWING 119 FOLLOWERS 196 LIKES 2

Following

## Arnstein Aassve

@Arnstein\_Aassve **FOLLOWS YOU**

I am the Dean for Undergraduate School at Bocconi University, where I am also Professor in Demography.

📍 Milan, Lombardy

🌐 [faculty.unibocconi.it/arnsteinaassve/](http://faculty.unibocconi.it/arnsteinaassve/)

📅 Joined March 2014

📄 Tweet to

💬 Message

👤 51 Followers you know

Tweets Tweets & replies Media



Arnstein Aassve Retweeted

marco albertini @madmakko · 5 Nov 2016

🗨️ @Tboeri for president

**Boeri : "Paternità obbligatoria per 15 giorni, e multe per chi non la fa"**

La proposta del presidente dell'Inps: «Fondamentale per un Paese come il nostro, in cui si penalizza in modo pesante la carriera delle madri». Poletti replica: vediamo i costi

🔄 2 ❤️ 2



Arnstein Aassve Retweeted

bruno arpino @brunoarpinoBCN · 27 Oct 2016

Higher prevalence of equal earners couples is found in regions where gender equality attitudes are more widespread. <https://doi.org/10.1016/j.jecl.2016.10.002>

Who to follow · Refresh · View all



Mark Girolami @MarkGiro... ✕

➤ Follow



Nial Friel @NialFriel ✕

➤ Follow



Julien JACQUES @jj\_stat ✕

➤ Follow

Find friends

Seattle Trends · Change



or tweets like this one:



## Tweets are stored in JSON format:

```
{ "created_at": "Wed Nov 07 04:16:18 +0000 2012",
  "id": 266031293945503744,
  "text": "Four more years. http://t.co/bAJE6Vom",
  "source": "web",
  "user": {
    "id": 813286,
    "name": "Barack Obama",
    "screen_name": "BarackObama",
    "location": "Washington, DC",
    "description": "This account is run by Organizing for Action staff.
      Tweets from the President are signed -bo.",
    "url": "http://t.co/8aJ56Jcemr",
    "protected": false,
    "followers_count": 40873124,
    "friends_count": 654580,
    "listed_count": 202495,
    "created_at": "Mon Mar 05 22:08:25 +0000 2007",
    "time_zone": "Eastern Time (US & Canada)",
    "statuses_count": 10687,
    "lang": "en" },
  "coordinates": null,
  "retweet_count": 783488,
  "favorite_count": 295026,
  "lang": "en"
}
```

## Tweets are stored in JSON format:

Key: value

```
{ "created_at": "Wed Nov 07 04:16:18 +0000 2012",  
  "id": 266031293945503744,  
  "text": "Four more years. http://t.co/bAJE6Vom",  
  "source": "web",  
  "user": {  
    "id": 813286,  
    "name": "Barack Obama",  
    "screen_name": "BarackObama",  
    "location": "Washington, DC",  
    "description": "This account is run by Organizing for Action staff.  
    Tweets from the President are signed -bo.",  
    "url": "http://t.co/8aJ56Jcemr",  
    "protected": false,  
    "followers_count": 40873124,  
    "friends_count": 654580,  
    "listed_count": 202495,  
    "created_at": "Mon Mar 05 22:08:25 +0000 2007",  
    "time_zone": "Eastern Time (US & Canada)",  
    "statuses_count": 10687,  
    "lang": "en" },  
  "coordinates": null,  
  "retweet_count": 783488,  
  "favorite_count": 295026,  
  "lang": "en"  
}
```

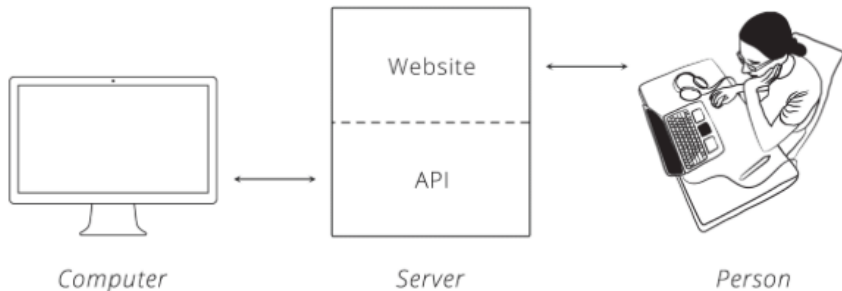
## Tweets are stored in JSON format:

```
{ "created_at": "Wed Nov 07 04:16:18 +0000 2012",  
  "id": 266031293945503744,  
  "text": "Four more years. http://t.co/bAJE6Vom",  
  "source": "web",  
  "user": {  
    "id": 813286,  
    "name": "Barack Obama",  
    "screen_name": "BarackObama",  
    "location": "Washington, DC",  
    "description": "This account is run by Organizing for Action staff.  
                  Tweets from the President are signed -bo.",  
    "url": "http://t.co/8aJ56Jcemr",  
    "protected": false,  
    "followers_count": 40873124,  
    "friends_count": 654580,  
    "listed_count": 202495,  
    "created_at": "Mon Mar 05 22:08:25 +0000 2007",  
    "time_zone": "Eastern Time (US & Canada)",  
    "statuses_count": 10687,  
    "lang": "en" },  
    "coordinates": null,  
    "retweet_count": 783488,  
    "favorite_count": 295026,  
    "lang": "en"  
  }  
}
```

Values can  
be lists

How do we access these data?


# How do we access these data?



Credit: Brian Cooksey

# Example of integrated Websites

# Example of integrated Websites



Find tacos, cheap dinner, Max's

Near Seattle, WA

Home About Me Write a Review Find Friends Messages Talk Events


## University of Washington

★★★★☆ 114 reviews Details

★ Write a Review

Add Photo Share Bookmark

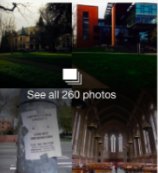
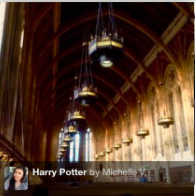

Colleges & Universities Edit



15th Ave NE  
Allen Library  
Map data ©2015 Google


**1400 NE Campus Pkwy**  
**Seattle, WA 98195**  
University District  
[Get Directions](#)  
(206) 543-2100  
[washington.edu](#)

[Edit](#)




See all 260 photos


1 Harry Potter by Michelle V.




"I've seen a handful of campuses and I have to say, I was incredibly impressed when I first visited **UW**." in 52 reviews





"Sitting under the cherry blossoms each spring on the **Quad**....like a snowstorm only with cherry blossom petals." in 16 reviews




"Fortunately, I came during early spring when the **cherry blossoms** were still there." in 9 reviews

 [Edit business info](#)

 [Work here? Claim this business](#)



Help fight food deserts.

DRINK  DO

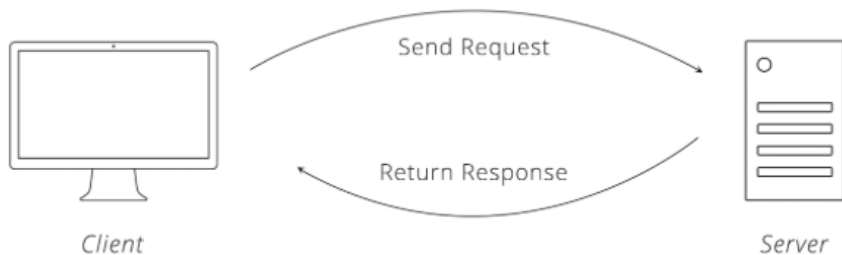


# University of Washington campus



Communication typically happens via HTTP

# Communication typically happens via HTTP



Credit: Brian Cooksey

# Authentication

# Authentication

- ▶ We have all registered for a website before (e.g., Facebook, Google, Dropbox, etc.):

# Authentication

- ▶ We have all registered for a website before (e.g., Facebook, Google, Dropbox, etc.):
- ▶ Logging-in with a username and password is an example of authentication

# Authentication

- ▶ We have all registered for a website before (e.g., Facebook, Google, Dropbox, etc.):
- ▶ Logging-in with a username and password is an example of authentication
- ▶ A different example: The developer of a game app can use the Dropbox API to let users store their saved games in the Dropbox cloud

# Authentication

- ▶ We have all registered for a website before (e.g., Facebook, Google, Dropbox, etc.):
- ▶ Logging-in with a username and password is an example of authentication
- ▶ A different example: The developer of a game app can use the Dropbox API to let users store their saved games in the Dropbox cloud
- ▶ Would you share your Dropbox username and password with the game developer?



# API Key Authentication

# API Key Authentication

- ▶ It is a technique to overcome the weakness of sharing credentials

# API Key Authentication

- ▶ It is a technique to overcome the weakness of sharing credentials
- ▶ The key is usually a long series of letters and numbers that is distinct from the account owner's login password

# API Key Authentication

- ▶ It is a technique to overcome the weakness of sharing credentials
- ▶ The key is usually a long series of letters and numbers that is distinct from the account owner's login password
- ▶ The owner gives the key to the application

# API Key Authentication

- ▶ It is a technique to overcome the weakness of sharing credentials
- ▶ The key is usually a long series of letters and numbers that is distinct from the account owner's login password
- ▶ The owner gives the key to the application
- ▶ The owner can restrict administrative privileges associated with the key

Let's look at some real-life examples and applications