

### 3 - Digital trace data

Diego Alburez-Gutierrez

MPIDR

European Doctoral School of Demography 2019-20

01/04/2020



MAX-PLANCK-INSTITUT  
FÜR DEMOGRAFISCHE  
FORSCHUNG

MAX PLANCK INSTITUTE  
FOR DEMOGRAPHIC  
RESEARCH

# Agenda

1. Q&A
2. Introduction to digital trace and marketing data
3. Example 1: Migration
4. Example 2: Internet users
5. Discussion

# Q&A

- ▶ Questions about on Exercise 1 from the final assignement
- ▶ Issues with Familinx data
- ▶ Other?

# Digital traces are incidental to our online presence

- ▶ Digital breadcrumbs are unavoidable
- ▶ Pre-GDPR, largely unchecked
- ▶ Marketing-led
- ▶ Not collected for social-scientific research

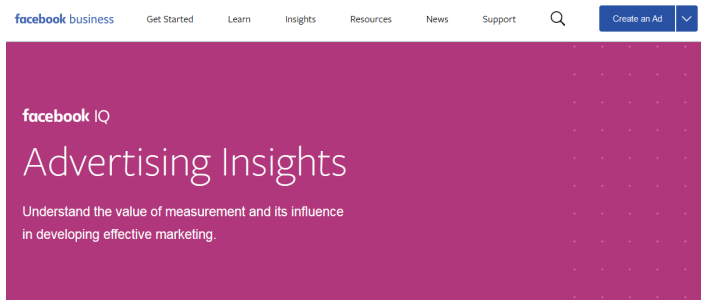
# Some data sources



- ▶ Marketing platforms
  - ▶ Facebook/Instagram/WhatsApp API
  - ▶ LinkedIn API
- ▶ Online platforms and communication
  - ▶ Twitter (API)
  - ▶ Google Trends
  - ▶ Email, IP address, mobile phones
- ▶ Internet of Things
  - ▶ Activity trackers and wearables medical devices
  - ▶ Wearable sensors (see Cito Cattuto)

# Facebook marketing platforms and APIs

- ▶ Sofia Gil's tutorial:  
[https://github.com/SofiaG1l/Using\\_Facebook\\_API](https://github.com/SofiaG1l/Using_Facebook_API)
- ▶ For python users:  
<https://github.com/carolcoimbra/facebook-ads>

# Using online marketing tools for demographic research

A screenshot of the Facebook Business Advertising Insights banner. The banner has a solid magenta background. On the left, the text 'facebook IQ' is in white, followed by 'Advertising Insights' in a larger white font. Below this, a smaller white line of text reads: 'Understand the value of measurement and its influence in developing effective marketing.' On the right side of the banner, there is a grid of small white 'x' marks arranged in approximately 10 rows and 10 columns.

facebook business   Get Started   Learn   Insights   Resources   News   Support      [Create an Ad](#) 

facebook IQ

## Advertising Insights

Understand the value of measurement and its influence in developing effective marketing.

# 'Audience estimates': FB users in Guatemala

The screenshot displays the Facebook Ads Manager interface. On the left, a sidebar contains navigation links: Campaign (Objective), Ad account (Create new), Ad set (Page, Audience, Placements, Budget & schedule), and Ad (Identity, Format, Media, Text). The main area is titled 'Diego Alburez (371284279)' and shows the ad set name '18+'. A map of Guatemala is centered on Guatemala City, with a 40 km radius. Below the map, the 'Locations' section includes a search bar and a 'Drop Pin' button. The 'Age' range is set to 18 to 65+, 'Gender' is set to All, and 'Languages' is set to 'Enter a language...'. The 'Include people who match' section shows 'Behaviours > Mobile Device User'. On the right, the 'Audience size' section indicates a potential reach of 3,700,000 people. The 'Estimated daily results' section shows a reach of 7.5K-22K. A note states: 'Estimate doesn't include Facebook Stories. Because Facebook Stories is a new placement being released gradually, audience and reach estimates aren't currently available. These estimates are based on the other placements that you've selected.'

Facebook Ads Manager interface showing audience estimates for Facebook users in Guatemala.

**Ad set name:** 18+

**Locations:** Guatemala, Guatemala City, Guatemala Department. + 40 km.

**Age:** 18 - 65+

**Gender:** All, Men, Women

**Languages:** Enter a language...

**Audience size:** Your audience is defined. Potential reach: 3,700,000 people.

**Estimated daily results:** Reach: 7.5K-22K.

**Estimate doesn't include Facebook Stories:** Because Facebook Stories is a new placement being released gradually, audience and reach estimates aren't currently available. These estimates are based on the other placements that you've selected.



# Male FB users, aged 18+ in Guatemala City

The screenshot displays the Facebook Ads Manager interface. The top navigation bar includes the Facebook logo, a hamburger menu, the text "Ads Manager", a search bar, and a user profile for "Diego". Below the navigation bar, the campaign is identified as "Diego Alburez (371284279)". The "Ad set name" is set to "18+", and there is a "Switch to Quick Creation" button.

The left sidebar contains a menu with the following items:

- Campaign
  - Objective
- Ad account
  - Create new
- Ad set
  - Page
  - Audience**
  - Placements
  - Budget & schedule
- Ad
  - Identity
  - Format
  - Media
  - Text

The main content area shows the "Audience" configuration for the "Ad set". It features a map of Guatemala with a blue circle centered on "Guatemala City, Guatemala Department". Below the map, the "Locations" section includes a "Drop Pin" button. The "Age" range is set to "18" to "65+", and the "Gender" is set to "Men". The "Languages" section has a text input field labeled "Enter a language...".

On the right side of the interface, there are several informational panels:

- Estimate doesn't include Facebook Stories**: A note stating that because Facebook Stories is a new placement, reach estimates are not currently available.
- Audience size**: A gauge showing the audience is defined, with a range from "Specific" to "Broad". Below the gauge, it states "Potential reach: 2,000,000 people".
- Estimated daily results**: A section showing "Reach" with a value of "6.4K-27K" and a corresponding progress bar.

At the bottom of the main content area, there is a section for "Include people who match" with a dropdown menu showing "Behaviours > Mobile Device User".

# Female FB users, aged 18+ in Guatemala City

The screenshot displays the Facebook Ads Manager interface. The top navigation bar includes the Facebook logo, a menu icon, the text "Ads Manager", a search bar, and a user profile for "Diego". Below the navigation bar, the left sidebar contains a "Campaign" section with a dropdown menu showing "Objective", "Ad account" with a "Create new" link, and "Ad set" with sub-options for "Page", "Audience" (highlighted in blue), "Placements", and "Budget & schedule". The main content area is titled "Ad set name" and shows "18+" as the selected age range. A map of Guatemala is displayed, with a pin placed in Guatemala City. Below the map, the "Locations" section shows "Guatemala" and "Guatemala City, Guatemala Department" with a "+ 40 km" radius. The "Age" range is set to "18" to "65+", and the "Gender" is set to "Women". The "Languages" section has a placeholder "Enter a language...". The "Include people who match" section is partially visible at the bottom. On the right side, a "Switch to Quick Creation" button is present. Below the map, there are three informational panels: "Estimate doesn't include Facebook Stories" with a note about the gradual release of Facebook Stories; "Audience size" showing a "Potential reach: 1,700,000 people" and a gauge for "Specific" and "Broad" audience definitions; and "Estimated daily results" showing a "Reach" of "8.1K-20K" with a corresponding bar chart.

Diego Alburez (371284279)

Ad set name 18+

Switch to Quick Creation

**Campaign**

- Objective

**Ad account**

- Create new

**Ad set**

- Page
- Audience**
- Placements
- Budget & schedule

**Ad**

- Identity
- Format
- Media
- Text

Close

Guatemala

Guatemala City, Guatemala Department

+ 40 km

Include Type to add more locations Browse

Locations

Tapachula, Quetzaltenango (Xela), Guatemala City, Antigua, Escuintla, Cobán, Zacapa, Chiquiriquí, Jutiapán, Atiquizaya, Acajutla, San Marcos

Drop Pin

Add locations in bulk

Age 18 - 65+

Gender All Men Women

Languages Enter a language...

Include people who match

Behavioral > Mobile Device > iOS

**Estimate doesn't include Facebook Stories**

Because Facebook Stories is a new placement being released gradually, audience and reach estimates aren't currently available. These estimates are based on the other placements that you've selected.

**Audience size**

Your audience is defined.

Potential reach: 1,700,000 people

**Estimated daily results**

Reach

**8.1K-20K**

The accuracy of estimates is based on factors such as past campaign data, the budget you've entered and market data. Numbers are provided to give you an idea of performance for your budget, but are only estimates and don't guarantee results.

# Substantive areas of interest

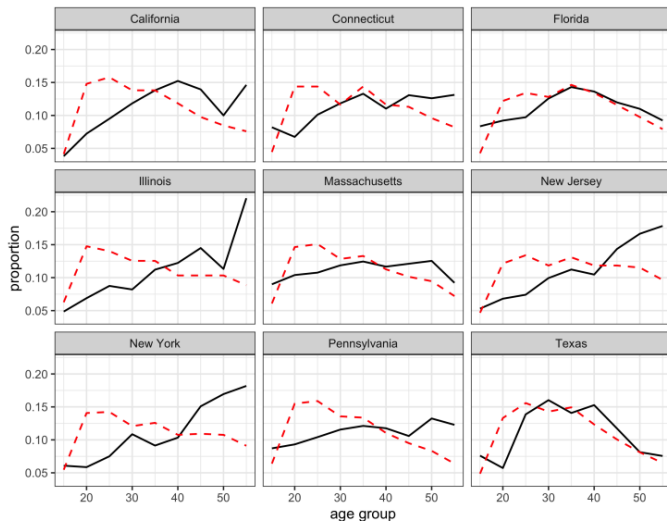
## Example 1: Migration

# Research at a glance

- ▶ Data
- ▶ Analysis
- ▶ Findings

Alexander, M., Polimis, K. and Zagheni, E. (2019), The Impact of Hurricane Maria on Out-migration from Puerto Rico: Evidence from Facebook Data. *Population and Development Review*, 45: 617-630.

# Sanity checks



Alexander, M., Polimis, K. and Zagheni, E. (2019), The Impact of Hurricane Maria on Out-migration from Puerto Rico: Evidence from Facebook Data. *Population and Development Review*, 45: 617-630.

# Population increase

Table 2: Estimated increase in Puerto Rican migrant stocks from October 2017 to January 2018. The 95% confidence intervals are shown in parentheses.

State (95% CI)	% Increase (95% CI)	Population Increase
Florida	21.6 (20.9, 22.3)	65433 (63342, 67525)
New York	11 (10.3, 11.7)	14477 (13584, 15371)
Pennsylvania	13.4 (12.7, 14.1)	13441 (12700, 14181)
Connecticut	14.7 (12.9, 16.5)	9402 (8244, 10560)
Massachusetts	10.1 (8.82, 11.4)	8957 (7824, 10090)
Texas	10.8 (10.4, 11.2)	5678 (5452, 5904)
Ohio	12.8 (12.2, 13.4)	3274 (3125, 3424)
Illinois	9.9 (9.15, 10.6)	2641 (2441, 2841)
Georgia	13.1 (12.4, 13.8)	2606 (2470, 2742)
New Jersey	2.9 (1.56, 4.24)	2282 (1228, 3336)
California	2.4 (1.86, 2.94)	573 (444, 702)

Alexander, M., Polimis, K. and Zagheni, E. (2019), The Impact of Hurricane Maria on Out-migration from Puerto Rico: Evidence from Facebook Data. *Population and Development Review*, 45: 617-630.

## Example 2: Digital use



# Summary

- ▶ Data
- ▶ Analysis
- ▶ Findings

# Gender gap

$$\text{Facebook Gender Gap Index} = \frac{\text{Female to male gender ratio of people with Facebook}}{\text{Female to Male gender ratio of the population}}$$

Fatehkia, M., Kashyap, R., and Weber, I. (2018). Using Facebook ad data to track the global digital gender gap. *World Development* 107:189–209.

# Use of digital trace data

- ▶ Strengths
  - ▶ Now-casting:  
<https://www.digitalgendergaps.org/data/?report=2020-03-02>
- ▶ Weaknesses

## Discussion

## Strengths of digital trace data

Study	Strengths	Weaknesses
Puerto Rico migration - Nowcasting \		
- Unique Digital access		

# Challenges going ahead

*Whoever you are. . . I've always depended on the kindness of strangers.*

— Blanche DuBois, *A Streetcar Named Desire*

1. Ensuring sustainable data access
2. Addressing systematic bias
3. Conducting ethical online research

# Privacy

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1):75–89.

# Homework

- ▶ Start with Exercise 2