Case Studies Using Social Media Marketing APIs

Ridhi Kashyap

Department of Sociology
Leverhulme Centre for Demographic Science
Nuffield College
University of Oxford

SICSS-Duke
June 21, 2022

Accessing Digital Trace Data

- Screen scraping
- Querying data from application programming interfaces (APIs)
- Archives or data sharing platforms (e.g. Google Trends).
- Data sharing agreements with data owners.

Screen Scraping

- The process of extracting data from web pages in an automated way
- Lots of data are available on the web – but not all of it is appropriate to use for research.

Considerations when Screen Scraping

- Read the website's Terms of Service: are you allowed to do this?
- Larger websites like Facebook, Instagram, NY Times do not allow these practices – but some provide structured access to data through APIs.
- Even if terms of service does not prohibit this – are the data sensitive? Could the use of the data harm in some way? Is their use ethical?

A Simple Example: Wikipedia Page

World Bank Group (2020) [edit]

Data of the World Bank Group for 2020.[3][4][5] The values in the World Bank Group tables are rounded. All calculations were done on raw data, therefore, due to the nuances of roun
places illusory inconsistencies of indicators arose, with a size of 0.01 year. Since this entry is prone to vandalism and political bias it is recommended to check the original source.

	Countries	all	male	female	gender gap	Δ 2019 all	Δ 2019 male	Δ 2019 female	Δ 2019 gen. gap
1	Japan	84.62	81.64	87.74	6.10	0.26	0.23	0.29	0.06
2	Singapore	83.74	81.50	86.10	4.60	0.15	0.10	0.20	0.10
3	South Korea	83.43	80.50	86.50	6.00	0.20	0.20	0.20	0.00
4	Norway	83.21	81.60	84.90	3.30	0.25	0.30	0.20	-0.10
5	Australia	83.20	81.20	85.30	4.10	0.30	0.30	0.30	0.00
6	Switzerland	83.10	81.10	85.20	4.10	-0.80	-1.00	-0.60	0.40
7	Iceland	83.07	81.70	84.50	2.80	-0.10	0.00	-0.20	-0.20
8	Israel	82.70	80.70	84.80	4.10	-0.10	-0.30	0.10	0.40
9	Malta	82.65	80.80	84.60	3.80	-0.20	-0.40	0.00	0.40
10	Sweden	82.41	80.70	84.20	3.50	-0.70	-0.80	-0.60	0.20
11	Italy	82.34	80.10	84.70	4.60	-1.15	-1.30	-1.00	0.30

Figure:
`https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy`

Wikipedia Page HTML

```
<!DOCTYPE html>
<html class="client-nojs" lang="en" dir="ltr">
<head>
<meta charset="UTF-8"/>
<title>List of countries by life expectancy - Wikipedia</title>
<script>document.documentElement.className="client-js";RLCONF={"wgBreakFrames":false,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""
"Life expectancy"],"wgPageContentLanguage":"en","wgPageContentModel":"wikitext","wgRelevantPageName":"List_of_countries_by_life_expectancy","wgRelev
RLSTATE={"ext.globalCssJs.user.styles":"ready","site.styles":"ready","user.styles":"ready","ext.globalCssJs.user":"ready","user":"ready","user.optio
"ext.gadget.switcher","ext.centralauth.centralautologin","ext.popups","ext.uls.compactlinks","ext.uls.interface","ext.growthExperiments.SuggestedEdi
<script>(RLQ=window.RLQ||[]).push(function(){mw.loader.implement("user.options@12s5i",function($,jQuery,require,module){mw.user.tokens.set({"patrolT
<link rel="stylesheet" href="/w/load.php?lang=en&amp;modules=ext.cite.styles%7Cext.uls.interlanguage%7Cext.visualEditor.desktopArticleTarget.noscrip
<script async="" src="/w/load.php?lang=en&amp;modules=startup&amp;only=scripts&amp;raw=1&amp;skin=vector"></script>
<meta name="ResourceLoaderDynamicStyles" content=""/>
<link rel="stylesheet" href="/w/load.php?lang=en&amp;modules=site.styles&amp;only=styles&amp;skin=vector"/>
<meta name="generator" content="MediaWiki 1.39.0-wmf.16"/>
<meta name="referrer" content="origin"/>
<meta name="referrer" content="origin-when-crossorigin"/>
<meta name="referrer" content="origin-when-cross-origin"/>
<meta name="format-detection" content="telephone=no"/>
<meta property="og:image" content="https://upload.wikimedia.org/wikipedia/commons/thumb/5/52/Comparison_gender_life_expectancy_WHO.svg/1200px-Compar
<meta property="og:image:width" content="1200"/>
<meta property="og:image:height" content="954"/>
<meta property="og:image" content="https://upload.wikimedia.org/wikipedia/commons/thumb/5/52/Comparison_gender_life_expectancy_WHO.svg/800px-Compari
<meta property="og:image:width" content="800"/>
<meta property="og:image:height" content="636"/>
<meta property="og:image" content="https://upload.wikimedia.org/wikipedia/commons/thumb/5/52/Comparison_gender_life_expectancy_WHO.svg/640px-Compari
<meta property="og:image:width" content="640"/>
<meta property="og:image:height" content="509"/>
<meta property="og:title" content="List of countries by life expectancy - Wikipedia"/>
<meta property="og:type" content="website"/>
<link rel="preconnect" href="//upload.wikimedia.org"/>
<link rel="alternate" media="only screen and (max-width: 720px)" href="//en.m.wikipedia.org/wiki/List_of_countries_by_life_expectancy"/>
<link rel="alternate" type="application/x-wiki" title="Edit this page" href="/w/index.php?title=List_of_countries_by_life_expectancy&amp;action=edit
<link rel="apple-touch-icon" href="/static/apple-touch/wikipedia.png"/>
<link rel="shortcut icon" href="/static/favicon/wikipedia.ico"/>
<link rel="search" type="application/opensearchdescription+xml" href="/w/opensearch_desc.php" title="Wikipedia (en)"/>
<link rel="EditURI" type="application/rsd+xml" href="//en.wikipedia.org/w/api.php?action=rsd"/>
<link rel="license" href="https://creativecommons.org/licenses/by-sa/3.0/"/>
<link rel="canonical" href="https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy"/>
<link rel="dns-prefetch" href="//meta.wikimedia.org" />
<link rel="dns-prefetch" href="//login.wikimedia.org"/>
</head>
<body class="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-subject mw-editable page-List_of_countries_by_life_expectancy rootpage-List_of_coun
<div id="mw-head-base" class="noprint"></div>
<div id="content" class="mw-body" role="main">
    <a id="top"></a>
```

A Simple Example: Wikipedia Page

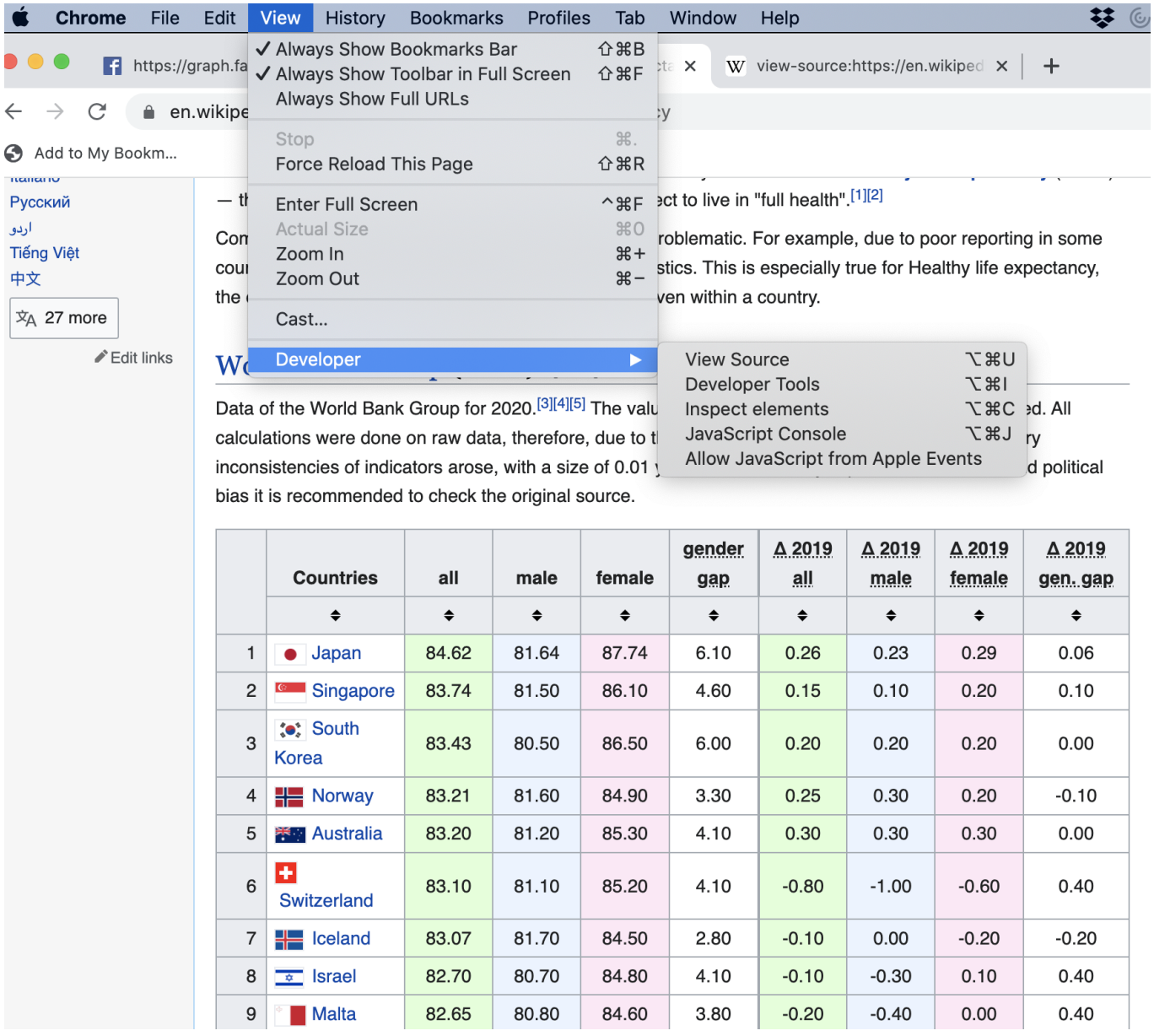
```
library(rvest)
le_wiki <- read_html("https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy")
le_table <- html_node(le_wiki, xpath = '//*[@id="mw-content-text"]/div/table[1]')
le_rankings <- html_table(le_table1)
head(le_rankings)

```

A tibble: 6 × 9

Countries <chr>	all <dbl>	male <dbl>	female <dbl>	gendergap <dbl>	Δ 2019all <dbl>	Δ 2019male <dbl>	Δ 2019female <dbl>
	NA	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
Japan	84.62	81.64	87.74	6.1	0.26	0.23	0.29
Singapore	83.74	81.50	86.10	4.6	0.15	0.10	0.20
South Korea	83.43	80.50	86.50	6.0	0.20	0.20	0.20
Norway	83.21	81.60	84.90	3.3	0.25	0.30	0.20
Australia	83.20	81.20	85.30	4.1	0.30	0.30	0.30

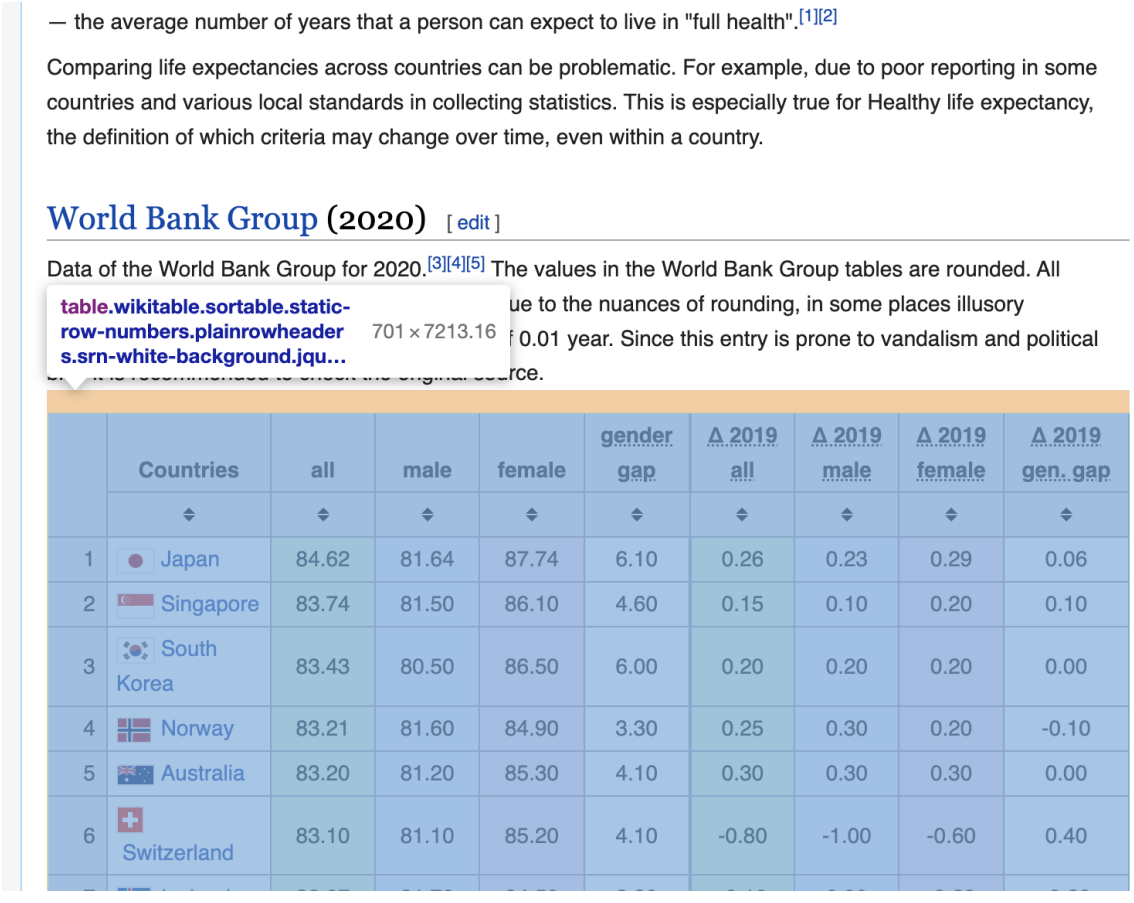
6 rows | 1–8 of 9 columns

A Simple Example: Wikipedia Page



	Countries	all	male	female	gender gap	Δ 2019 all	Δ 2019 male	Δ 2019 female	Δ 2019 gen. gap
1	Japan	84.62	81.64	87.74	6.10	0.26	0.23	0.29	0.06
2	Singapore	83.74	81.50	86.10	4.60	0.15	0.10	0.20	0.10
3	South Korea	83.43	80.50	86.50	6.00	0.20	0.20	0.20	0.00
4	Norway	83.21	81.60	84.90	3.30	0.25	0.30	0.20	-0.10
5	Australia	83.20	81.20	85.30	4.10	0.30	0.30	0.30	0.00
6	Switzerland	83.10	81.10	85.20	4.10	-0.80	-1.00	-0.60	0.40
7	Iceland	83.07	81.70	84.50	2.80	-0.10	0.00	-0.20	-0.20
8	Israel	82.70	80.70	84.80	4.10	-0.10	-0.30	0.10	0.40
9	Malta	82.65	80.80	84.60	3.80	-0.20	-0.40	0.00	0.40

A Simple Example: Wikipedia Page



— the average number of years that a person can expect to live in "full health".[1][2]

Comparing life expectancies across countries can be problematic. For example, due to poor reporting in some countries and various local standards in collecting statistics. This is especially true for Healthy life expectancy, the definition of which criteria may change over time, even within a country.

World Bank Group (2020) [edit]

Data of the World Bank Group for 2020.[3][4][5] The values in the World Bank Group tables are rounded. All
ue to the nuances of rounding, in some places illusory
0.01 year. Since this entry is prone to vandalism and political
rce.

table.wikitable.sortable.static-row-numbers.plainrowheaders.srn-white-background.jqu... 701 × 7213.16

	Countries	all	male	female	gender gap	Δ 2019 all	Δ 2019 male	Δ 2019 female	Δ 2019 gen. gap
1	Japan	84.62	81.64	87.74	6.10	0.26	0.23	0.29	0.06
2	Singapore	83.74	81.50	86.10	4.60	0.15	0.10	0.20	0.10
3	South Korea	83.43	80.50	86.50	6.00	0.20	0.20	0.20	0.00
4	Norway	83.21	81.60	84.90	3.30	0.25	0.30	0.20	-0.10
5	Australia	83.20	81.20	85.30	4.10	0.30	0.30	0.30	0.00
6	Switzerland	83.10	81.10	85.20	4.10	-0.80	-1.00	-0.60	0.40

Some Limits

- This was a simple example. Often web pages are complex with many different elements.
- This yields messy data.
- Screen scraping can be frustrating or unfeasible. What then?
  - For complex web pages, crowd worker platforms (e.g. Mechanical Turk) could be an option.
  - Some web data can be accessed via APIs.

What is an API?

- Application programming interfaces or APIs are a software intermediary that allows two applications to talk to each other.
- Web APIs allow one computer (a client) to ask another computer (a server) for some resource over the internet.
- APIs provide a structured way to access data that are stored in databases that are continuously updated.

Application Programming Interface

- Modern APIs adhere to standards, that make data exchange programmatically accessible, safe and structured
- Contrast with web scraping.
- Two important concepts when using APIs
  - Credentialling
  - Rate limiting

API Directory

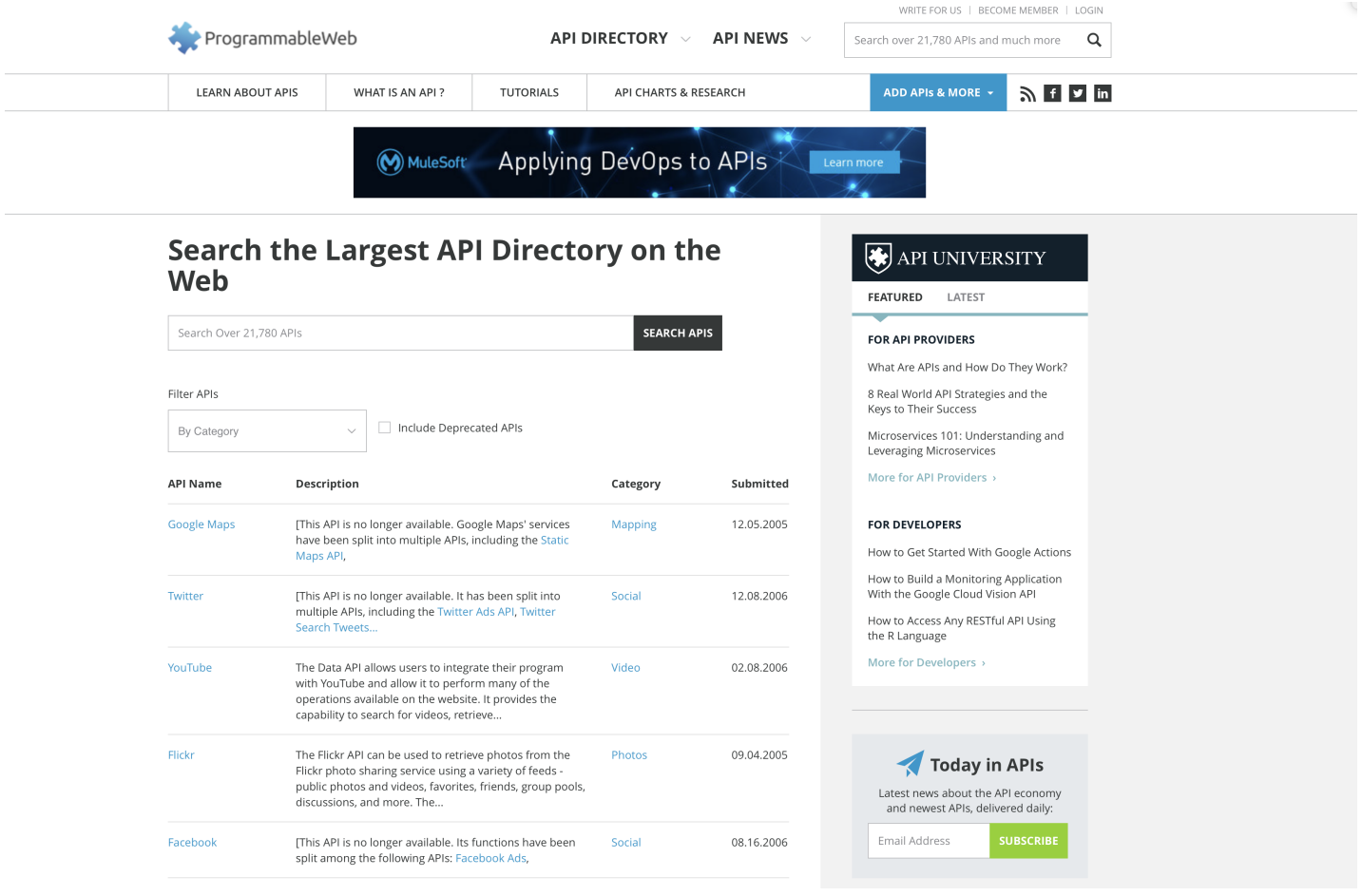


Figure: `https://www.programmableweb.com/apis/directory`

httr in R

- The `httr` package in R is useful to work with APIs
- You can make a request (to a url) and get a response.
- Response contains a status, header and body.

Social Media Ad Audience Estimates

- **Digital census**: How many users of 'x' characteristics (age, gender, location, etc) are on a given platform?

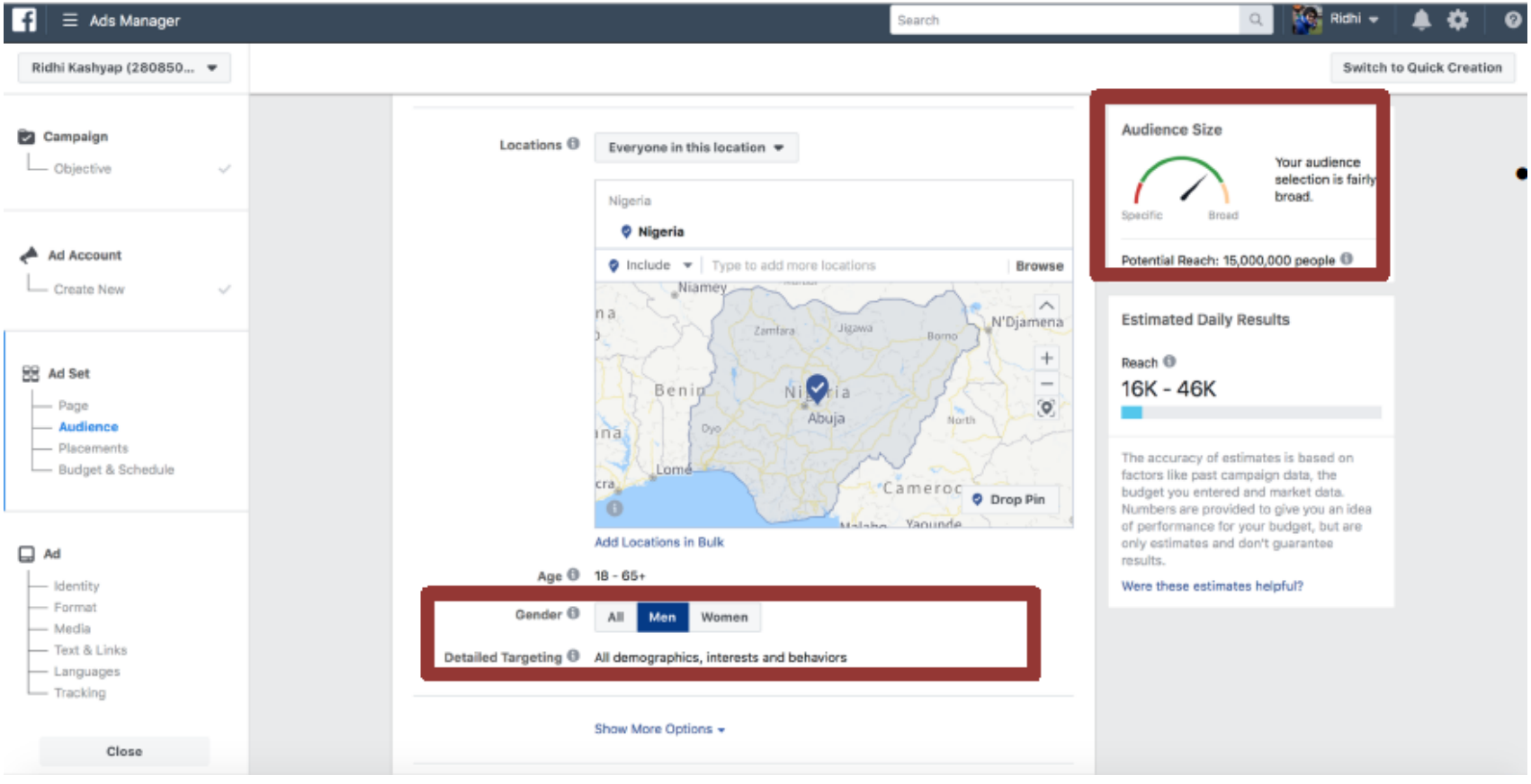


Figure: Facebook ads manager: `https://www.facebook.com/adsmanager`

Social Media Ad Audience Estimates

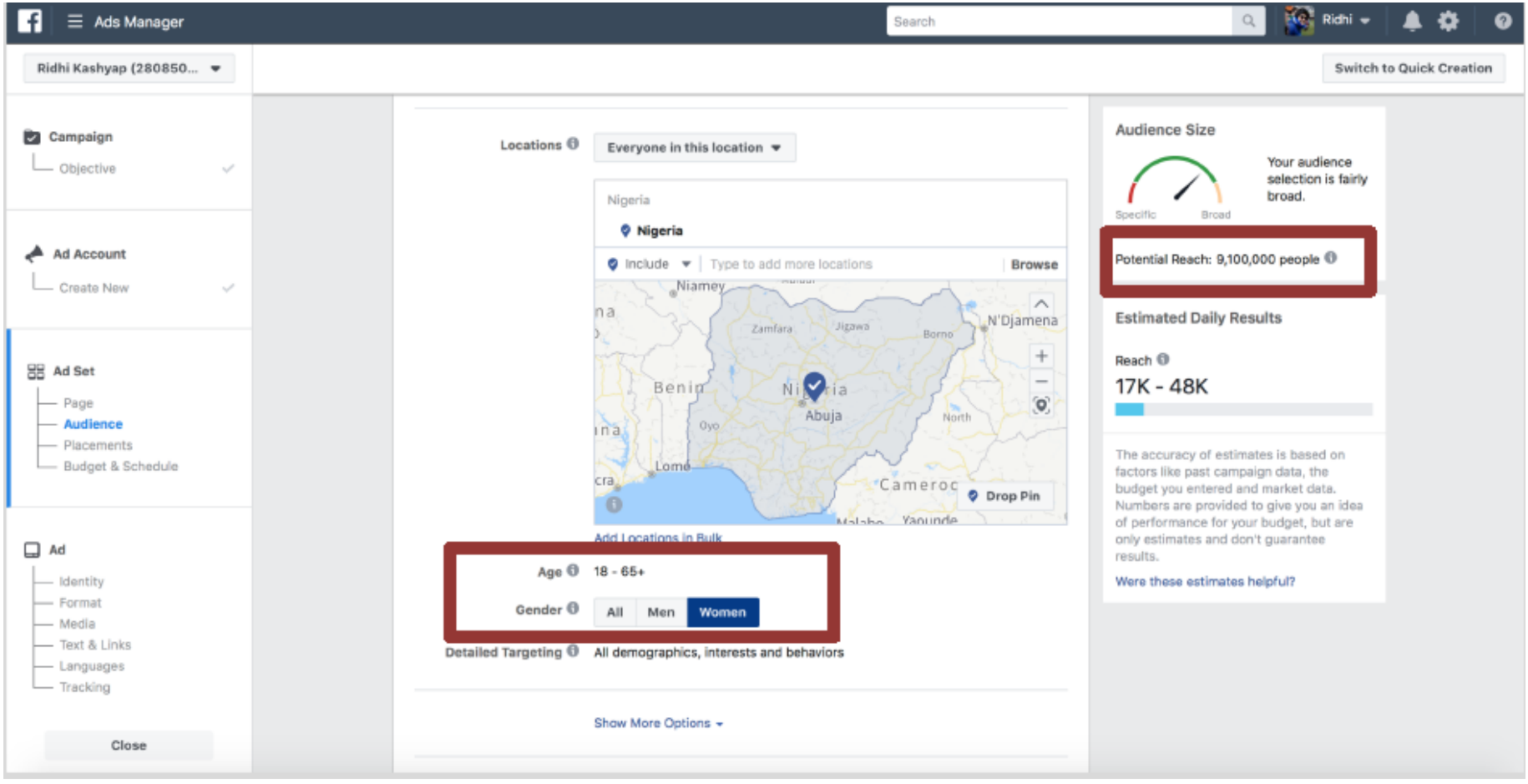


Figure: Facebook ads manager: `https://www.facebook.com/adsmanager`

Social Media Ad Audience Estimates: Updated MAU

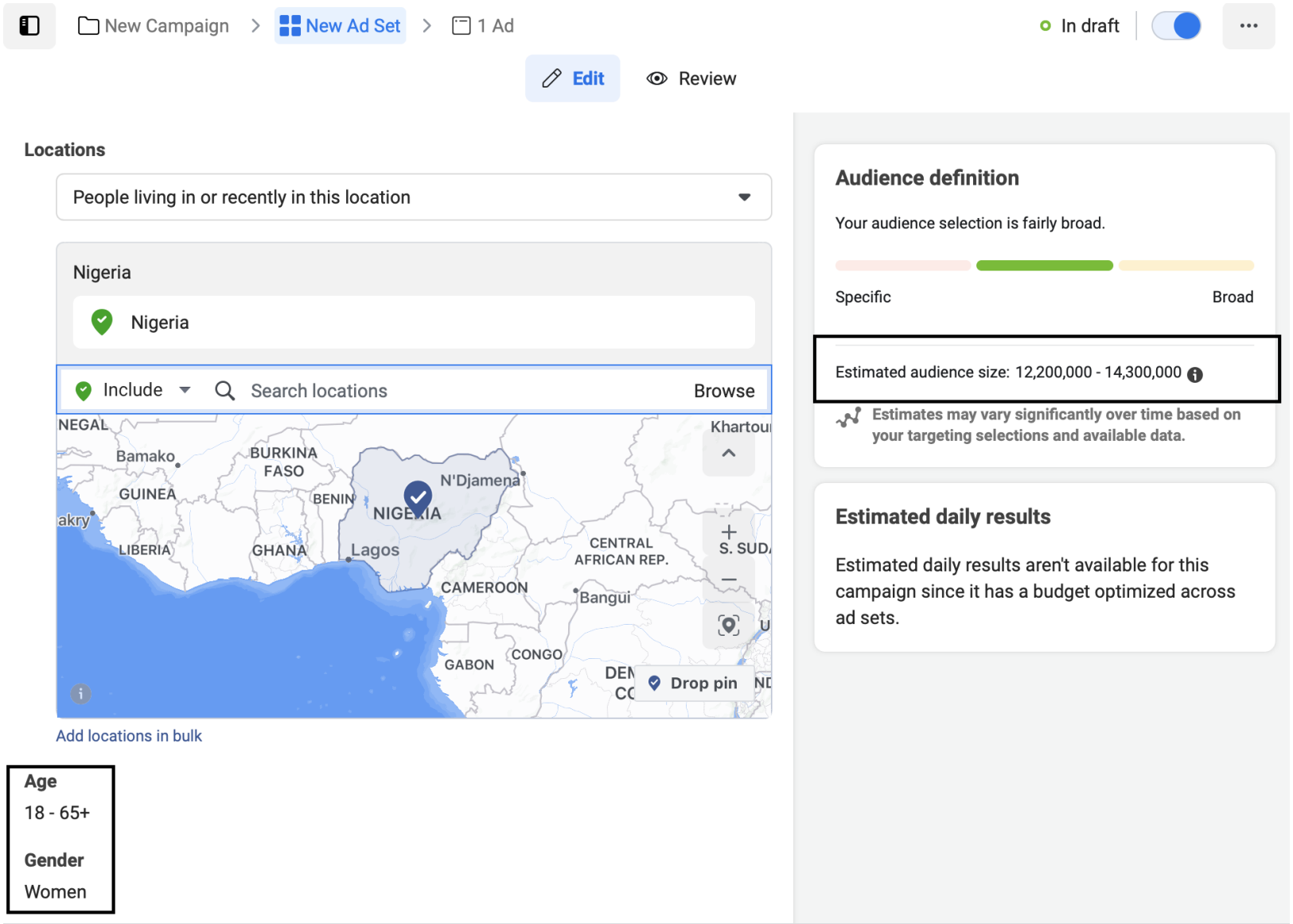


Figure: Facebook ads manager: `https://www.facebook.com/adsmanager`

Social Media Ad Audience Estimates

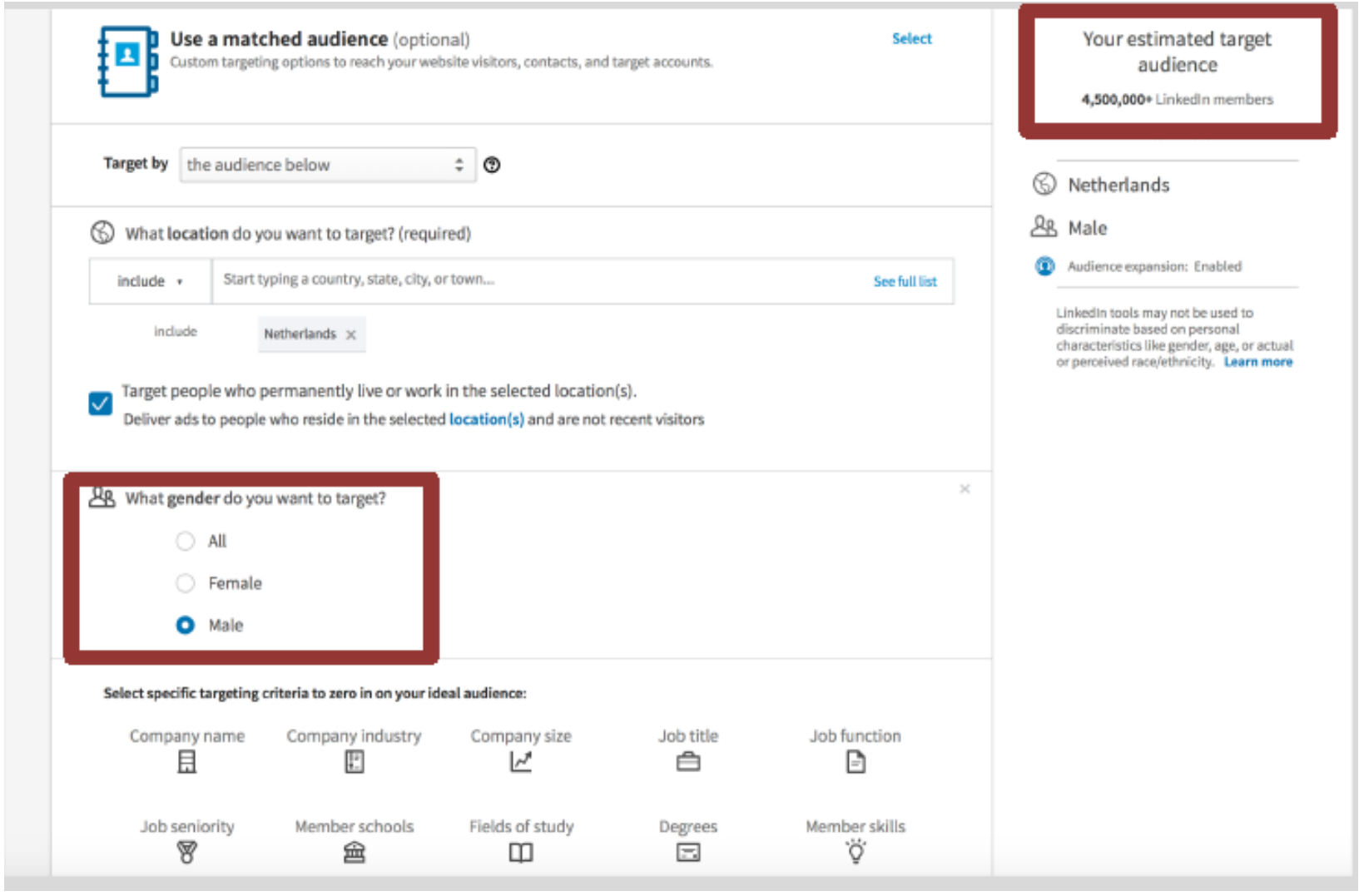
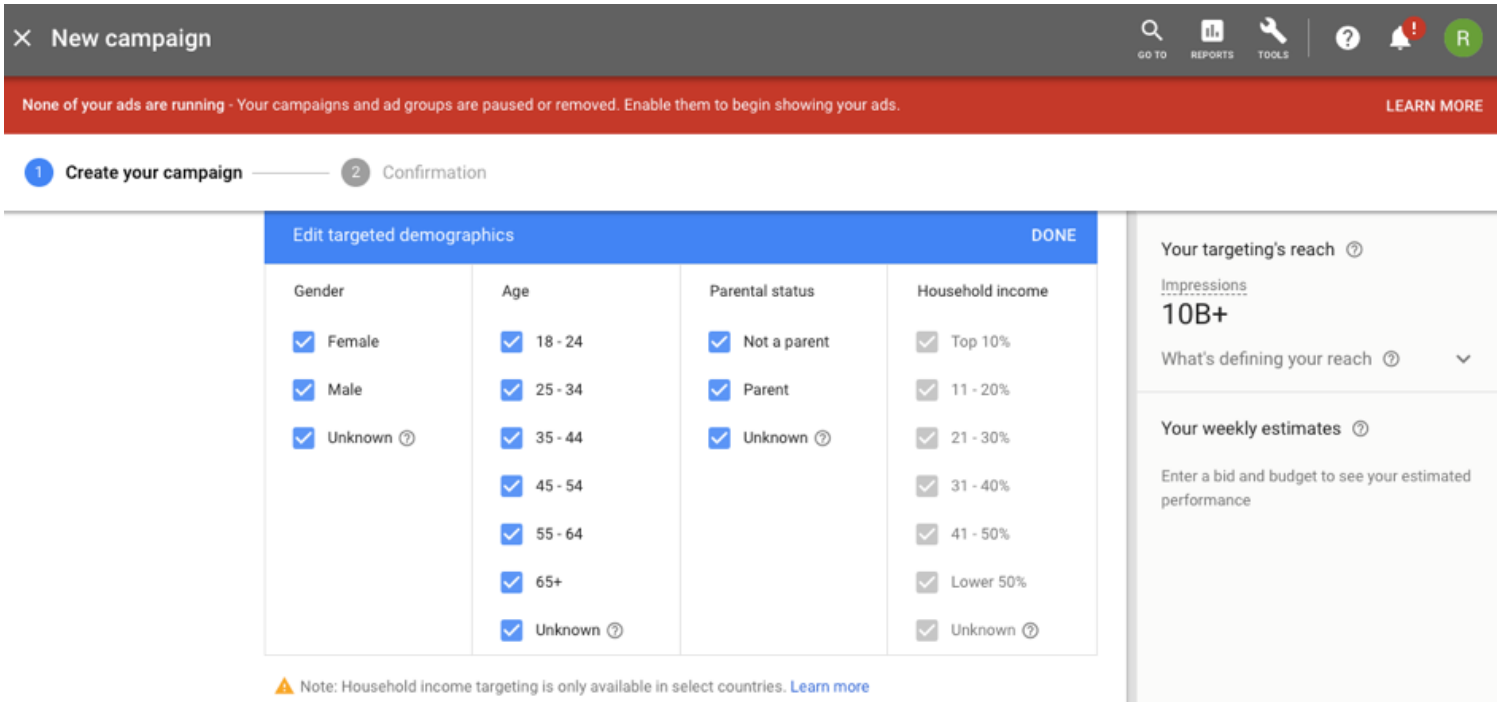


Figure: LinkedIn ads manager

AdWords


New campaign
GO TO
REPORTS
TOOLS
R
None of your ads are running - Your campaigns and ad groups are paused or removed. Enable them to begin showing your ads.
LEARN MORE
1 Create your campaign
2 Confirmation
Edit targeted demographics
DONE
Gender
Female
Male
Unknown
Age
18 - 24
25 - 34
35 - 44
45 - 54
55 - 64
65+
Unknown
Parental status
Not a parent
Parent
Unknown
Household income
Top 10%
11 - 20%
21 - 30%
31 - 40%
41 - 50%
Lower 50%
Unknown
Note: Household income targeting is only available in select countries. Learn more
Your targeting's reach
Impressions
10B+
What's defining your reach
Your weekly estimates
Enter a bid and budget to see your estimated performance

Underlying Data: Facebook Marketing API

- You need:
  - Facebook account
  - Marketing app with token and an ad account number ("act"): these are the credentials to make calls to the API
  - Steps for obtaining these credentials are at: `https://github.com/ridhi-kashyap/SICSS_Digital_Trace_2022/blob/main/Steps_for_creating_FB_Access_Token.pdf`

Facebook Marketing API

- We can programmatically make a query to the API to retrieve these ad audience estimates.
- For more information on how targeting specifications work – see `https://developers.facebook.com/docs/marketing-api/buying-api/targeting/`
- To search for available targeting options – see `https://developers.facebook.com/docs/marketing-api/targeting-search`

Facebook Marketing API

```r
#Loading required packages
library(readr)
library(dplyr)
library(jsonlite)
library(httr)

#Specify version of the API
version <- "v13.0"

#Specify your authentication/credentials,
#these should be saved separately in a private file

credential<-read.csv("credentials.csv",header = FALSE)

token <- credential$V1
act <- credential$V2

#We specify the basic URL as a string
Credentials <- paste0('https://graph.facebook.com/',version,'/act_',act,'/delivery_estimate?acces
s_token=',token)
```

- `version` refers to the version of the API (v 13.0 is current)
- `act` is the ad account number
- `token` is the access token
- Important to remember rate limiting when working with APIs and making multiple calls (Sys.sleep)

Facebook Marketing API: Basic Targeting Spec

```
targeting_spec_simple <- '{"geo_locations":{"countries":["GB"]}}'

query_char <- list(
  include_headers="false",
  method="get",
  optimization_goal="REACH",
  suppress_http_code=1,
  targeting_spec = targeting_spec_simple)
```

- Targeting specifications need to be specified in a JSON array – but we are treating it as a string here
- This is a call for one country's audience estimates.
- Other arguments for the query: we want the reach estimate, which is obtained via an HTTP get request

Facebook Marketing API: Requesting an Estimate

```
query_val1 <- GET(url = Credentials, query = query_char) %>% content(as="text",encoding =
"UTF-8") %>% fromJSON
query_val1<-query_val1$data
query_val1
#The query provides three counts -  1. estimated daily active users (dau)
#                                   two monthly active user
#                                   2. mau_upper_bound
#                                   3. mau_lower_bound

query_val1$estimate_dau
query_val1$estimate_mau_lower_bound
query_val1$estimate_mau_upper_bound
```

- We use GET from the `httr` package to obtain the estimates.
- These are returned as a JSON object.
- We extract them using the command from `fromJSON` from `JSONlite`

Facebook Marketing API: Output JSON


```
{
   "data": [
      {
         "daily_outcomes_curve": [
            {
               "spend": 0,
               "reach": 0,
               "impressions": 0,
               "actions": 0
            }
         ],
         "estimate_dau": 46394479,
         "estimate_mau_lower_bound": 45500000,
         "estimate_mau_upper_bound": 53600000,
         "estimate_ready": true
      }
   ]
}
```


- If we put the query into a browser this is what we get
- This is a JSON object.

Facebook Marketing API: Output JSON

```
query_val1 <- GET(url = Credentials, query = query_char) %>% content(as="text",encoding = "UTF-8") %>% fromJSON
query_val1<-query_val1$data
query_val1
```

```
##   daily_outcomes_curve estimate_dau estimate_mau_lower_bound
## 1           0, 0, 0, 0     46394479                 45500000
##   estimate_mau_upper_bound estimate_ready
## 1                 53600000           TRUE
```

```
#The query provides three counts -  1. estimated daily active users (dau)
#                                   two monthly active user
#                                   2. mau_upper_bound
#                                   3. mau_lower_bound

query_val1$estimate_dau
```

```
## [1] 46394479
```

```
query_val1$estimate_mau_lower_bound
```

```
## [1] 45500000
```

```
query_val1$estimate_mau_upper_bound
```

```
## [1] 53600000
```

Facebook Marketing API: Including Age as a Targeting Param

```
    target_query <- paste0('{"age_min":',age_min,
                  ',"age_max":',age_max,
                  ',"genders":[',genders,']',
                  ',"geo_locations":{"countries":[',countries,'],"location_types":["home","recent
"]}}')}
```

- We can add age through the age (min, max) arguments.
- For making multiple calls, e.g. multiple age groups and countries, it is helpful to write a function and loop.
- See example at: `https://github.com/ridhi-kashyap/SICSS_Digital_Trace_2022`

Facebook Marketing Search API: Requesting Available Countries

```
search_url <- "https://graph.facebook.com/v13.0/search"
country_search <-GET(url = search_url,
  query=list(
    type='adgeolocation',
    location_types='country',
    access_token=token,
    limit=1000)) %>%content(as="text", encoding="UTF-8")%>%fromJSON
```

- We can use the Facebook Marketing Search API to get obtain tables of targeting specifications (e.g. demographics, behaviours).
- The URL for this is `https://graph.facebook.com/v13.0/search`

Applications

- Predicting gender gap indicators:
  - Internet and mobile access gender gaps and digital skills gender gaps (using Facebook, Google AdWords)
  - Professional gender gaps (using LinkedIn)
- Monitoring population displacement in Ukraine during the war

Nowcasting Digital Gender Gaps

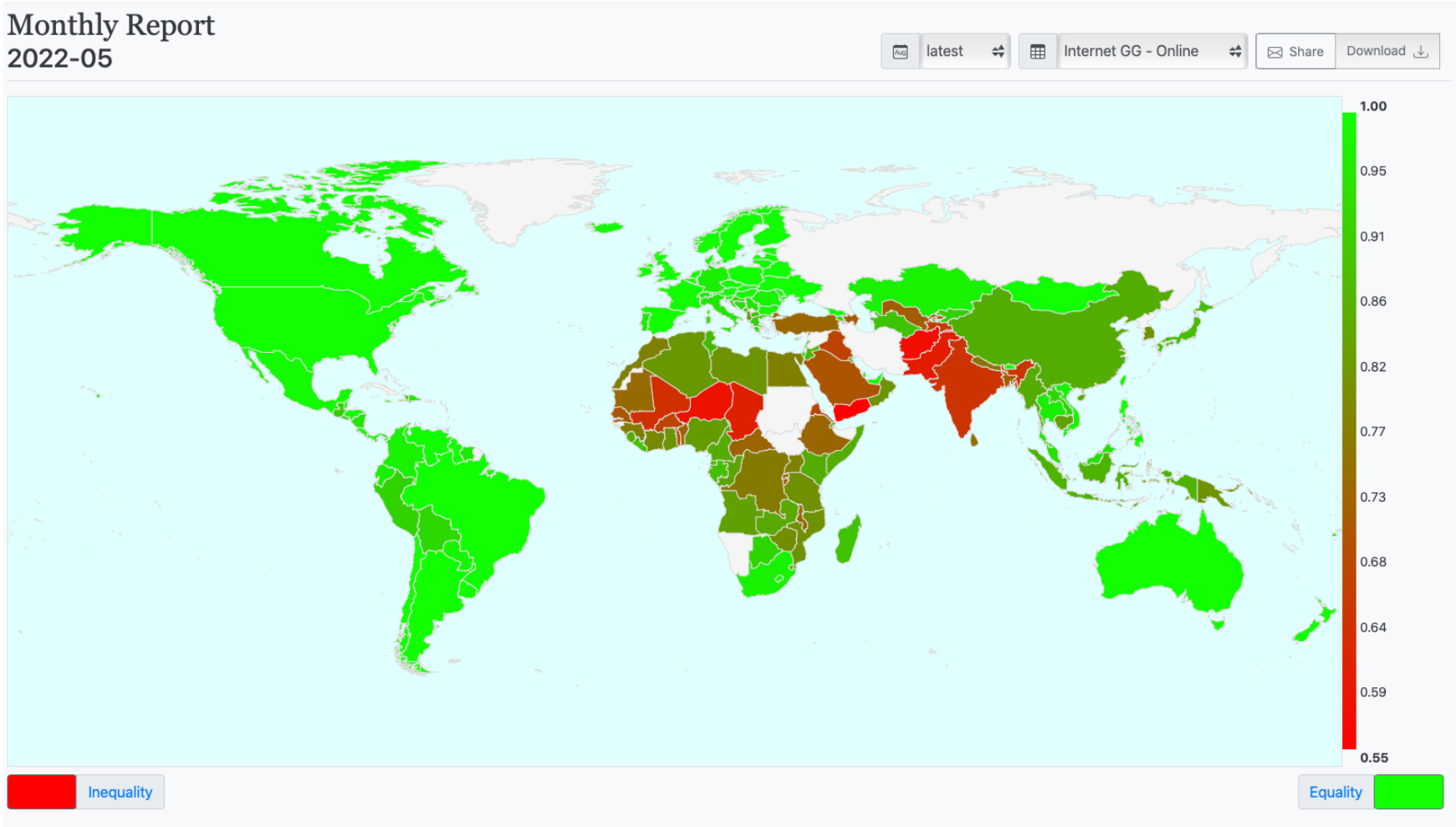
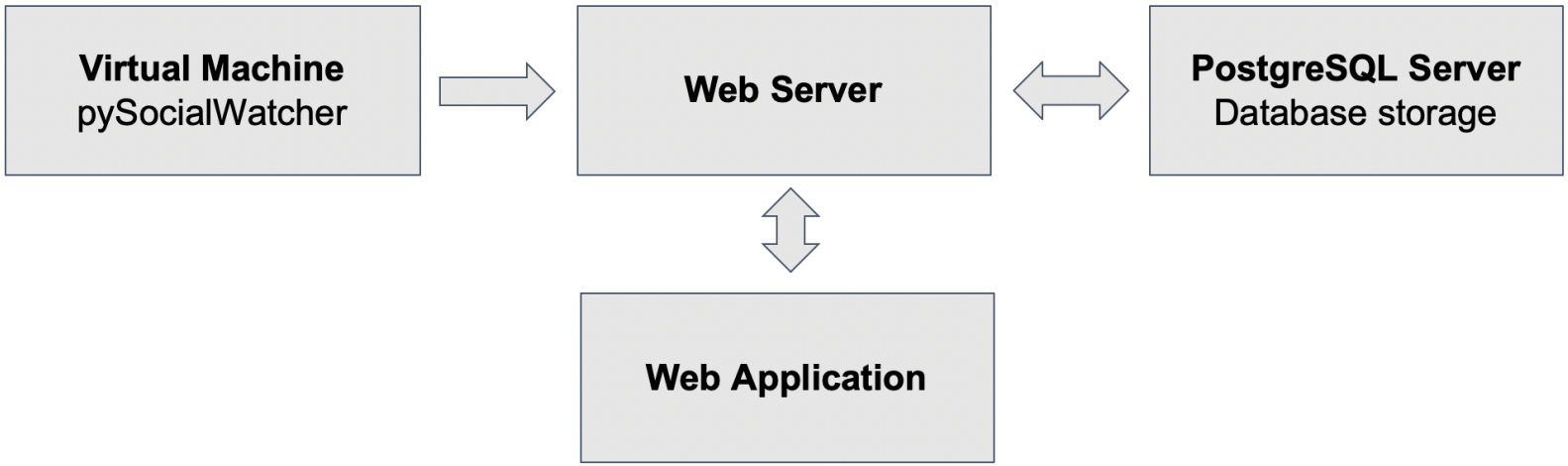


Figure: Gender gaps in Internet access predicted using Facebook gender gap index available at `www.digitalgendergaps.org`.

[1] Fatehkia, Masoomali, Ridhi Kashyap, and Ingmar Weber. "Using Facebook ad data to track the global digital gender gap." *World Development* 107 (2018): 189-209.

Backend Setup



LinkedIn Gender Gap Index

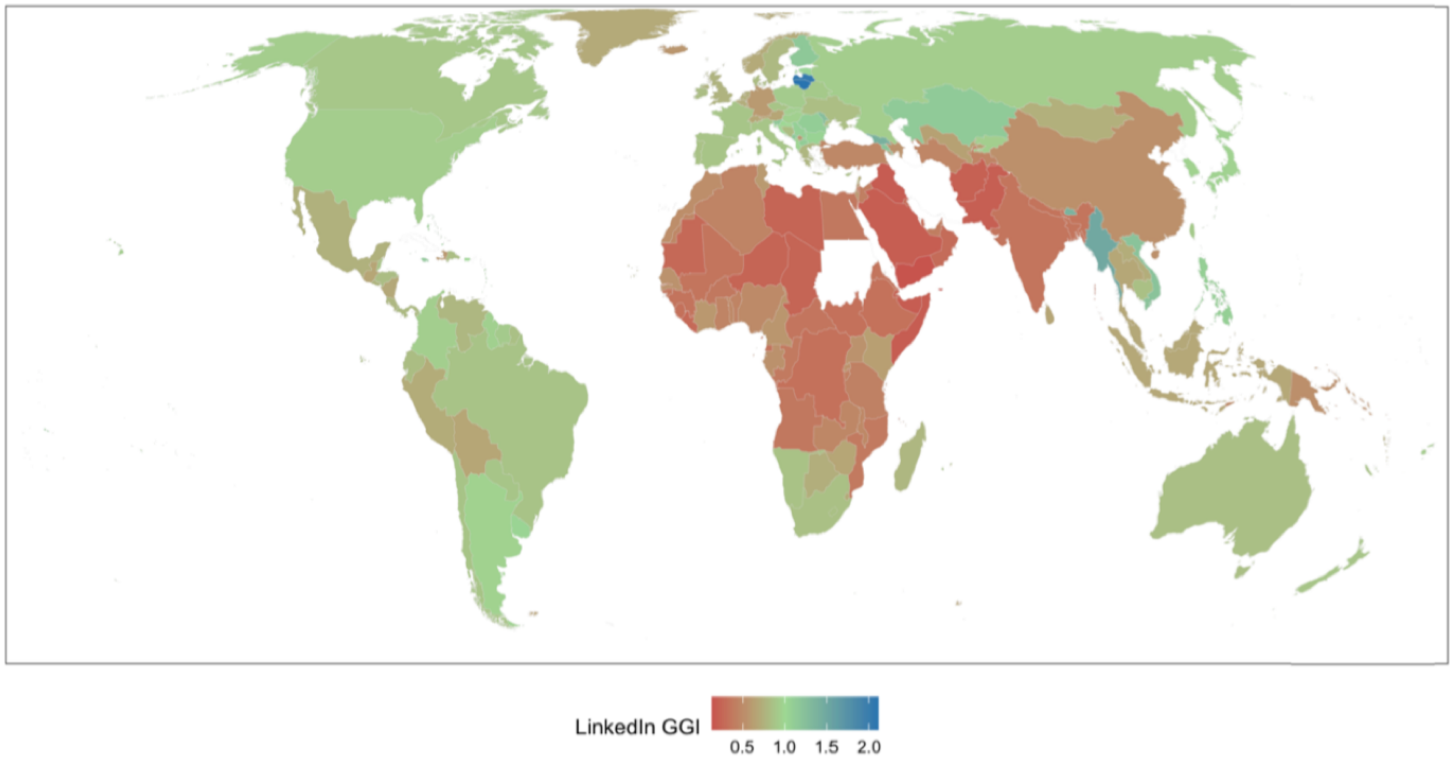


Figure: LinkedIn gender gap index (GGI) (female-to-male ratio of LinkedIn users), across the world

[1] Kashyap, Ridhi and Florianne Verkroost "Analysing global professional gender gaps using LinkedIn advertising data." *EPJ Data Science* 10(1): 39

Professional Gender Gaps

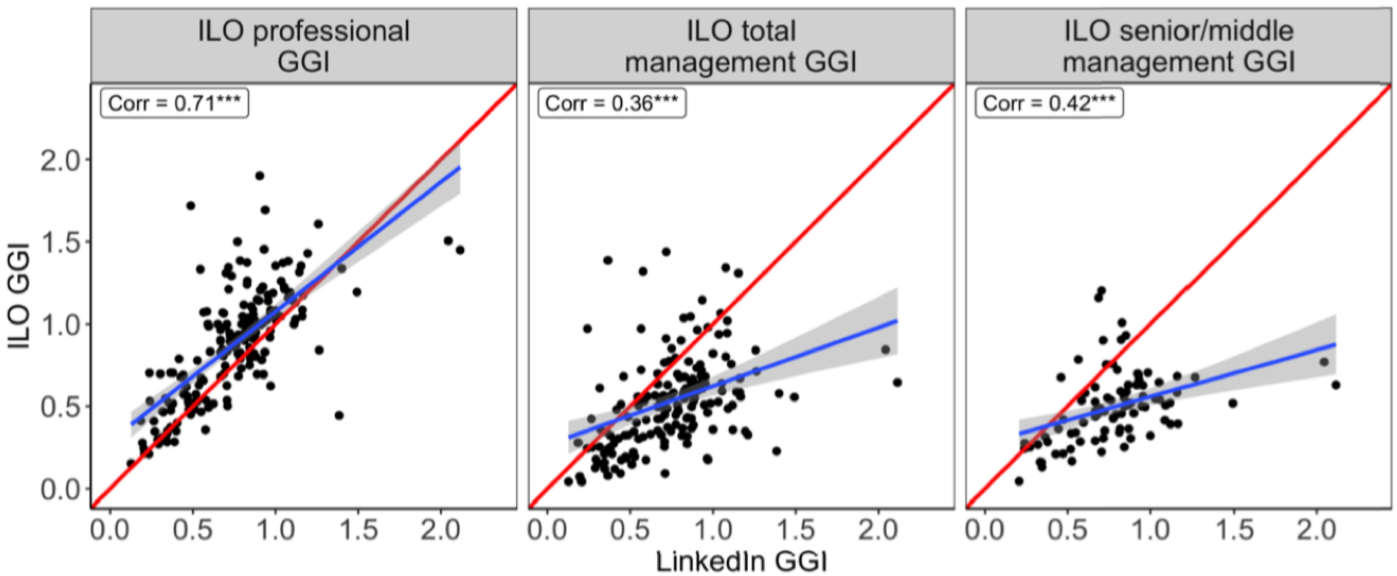
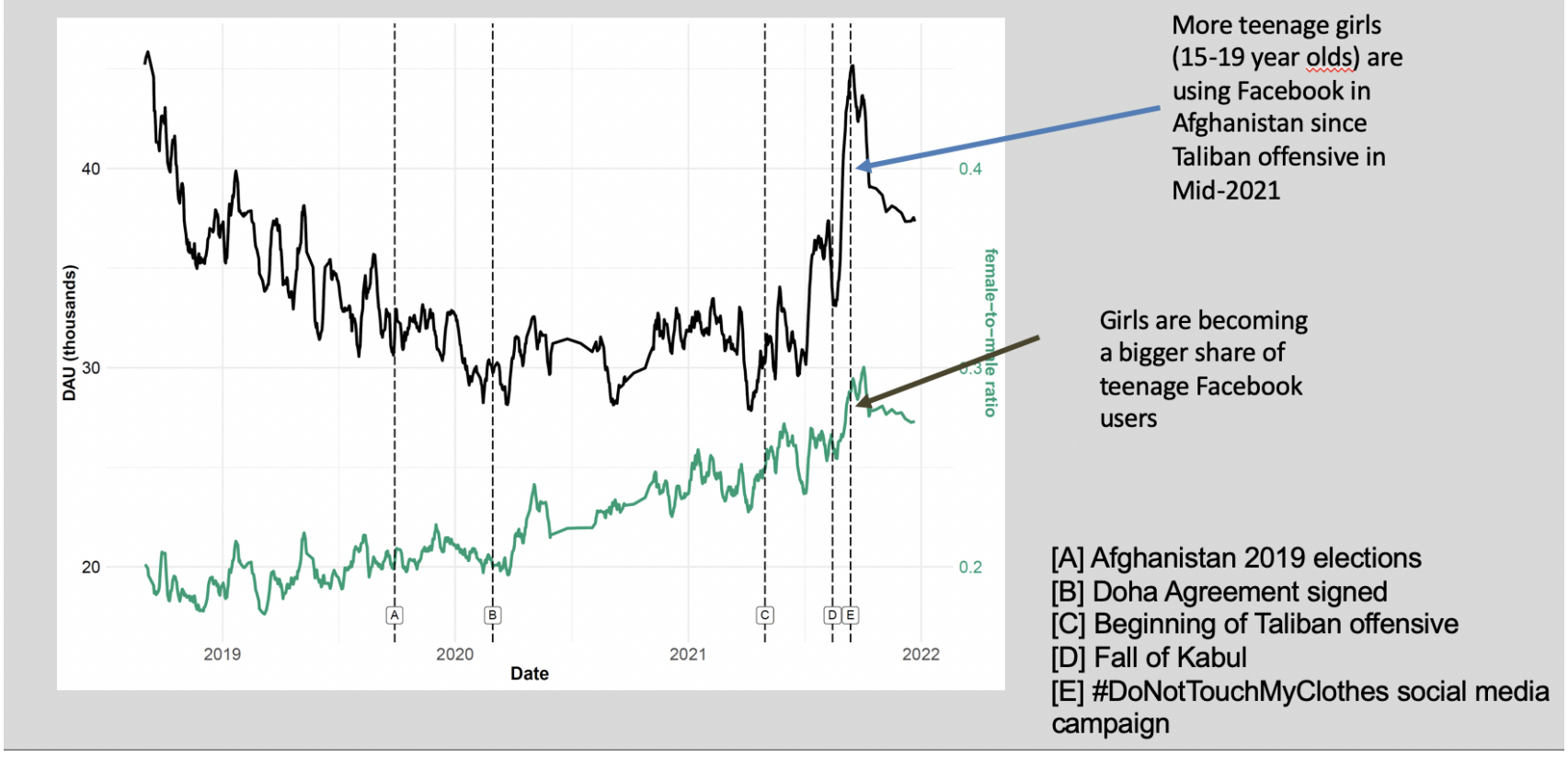


Figure: LinkedIn Gender Gap Index (GGI) and its correlations with different International Labour Organization (ILO) Gender Gaps

[1] Kashyap, Ridhi and Florianne Verkroost "Analysing global professional gender gaps using LinkedIn advertising data." *EPJ Data Science* 10(1): 39

Further applications

- Tracking changes during crises, e.g. Afghanistan



Ukraine

February 24, 2022

Objectives

- Can we use FB marketing API data to monitor population displacement within Ukraine during the war?
- Produce daily sub-national population estimates inside Ukraine disaggregated by age and sex.
- Help fill critical data gaps to inform humanitarian response efforts.

Socarxiv Preprint

Ukraine Crisis: Monitoring population displacement through social media activity

2022-06-06

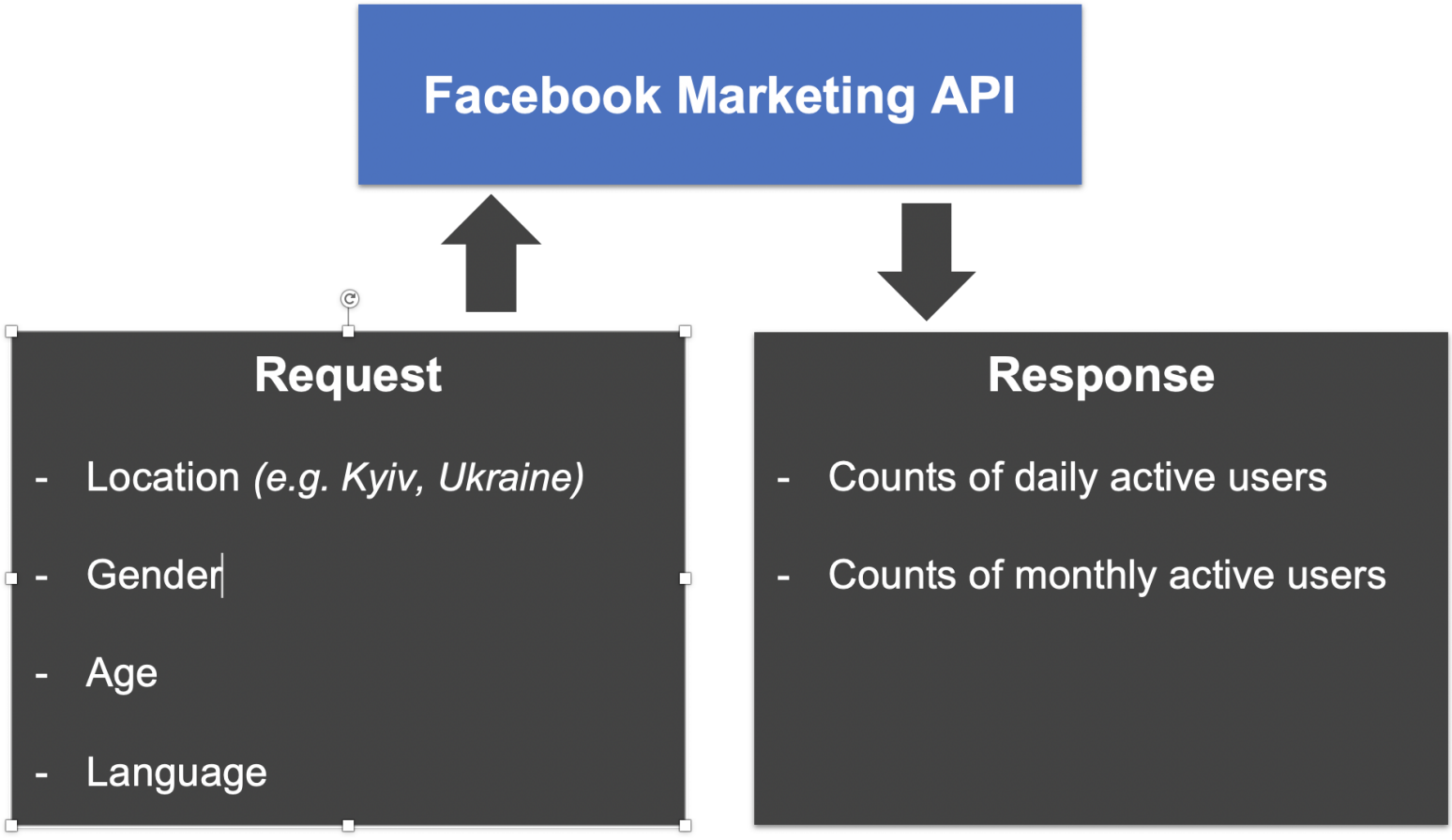
Douglas R. Leasure[1,3,6,*], Ridhi Kashyap[1,3], Francesco Rampazzo[2,1,3], Benjamin Elbers[3,1], Claire Dooley[4,6], Ingmar Weber[5], Masoomali Fatehkia[5], Maksym Bondarenko[6], Mark Verhagen[1], Arun Frey[1], Jiani Yan[1], Evelina T. Akimova[1], Robert Trigwell[7], Brian McDonald[7], Mohamed Bakr[7], Alessandro Sorichetta[6], Andrew J. Tatem[6], Melinda C. Mills[1,3]

[1] Leverhulme Centre for Demographic Science, Department of Sociology, University of Oxford
[2] Saïd Business School, University of Oxford
[3] Nuffield College, University of Oxford
[4] Department of Population Health, London School of Hygiene and Tropical Medicine
[5] Qatar Computing Research Institute, Hamad bin Khalifa University
[6] WorldPop, University of Southampton
[7] International Organization for Migration, United Nations
[*] douglas.leasure@sociology.ox.ac.uk

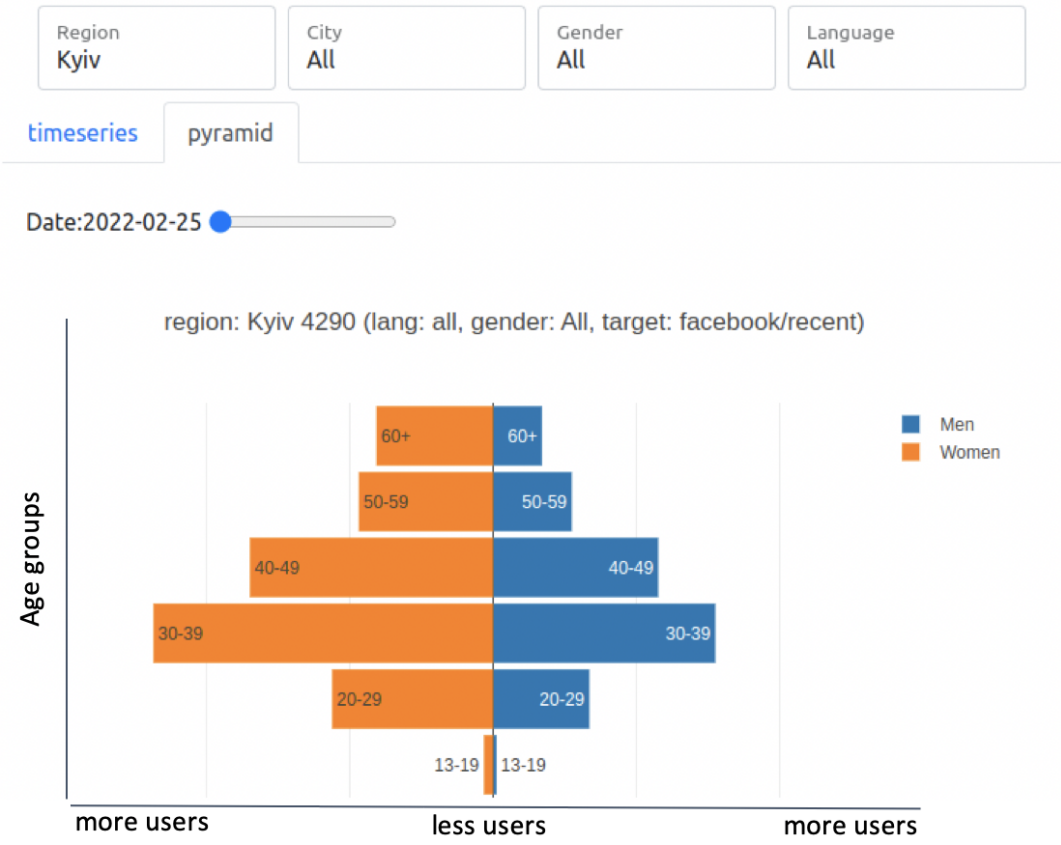
> **Note:** This analysis is a rapid-response effort that has not yet undergone peer-review. All results are provisional and should be interpreted with caution. Version updates will be provided as potential issues are identified, methods are improved, or new data become available.

Figure: `https://doi.org/10.31235/osf.io/6j9wq`

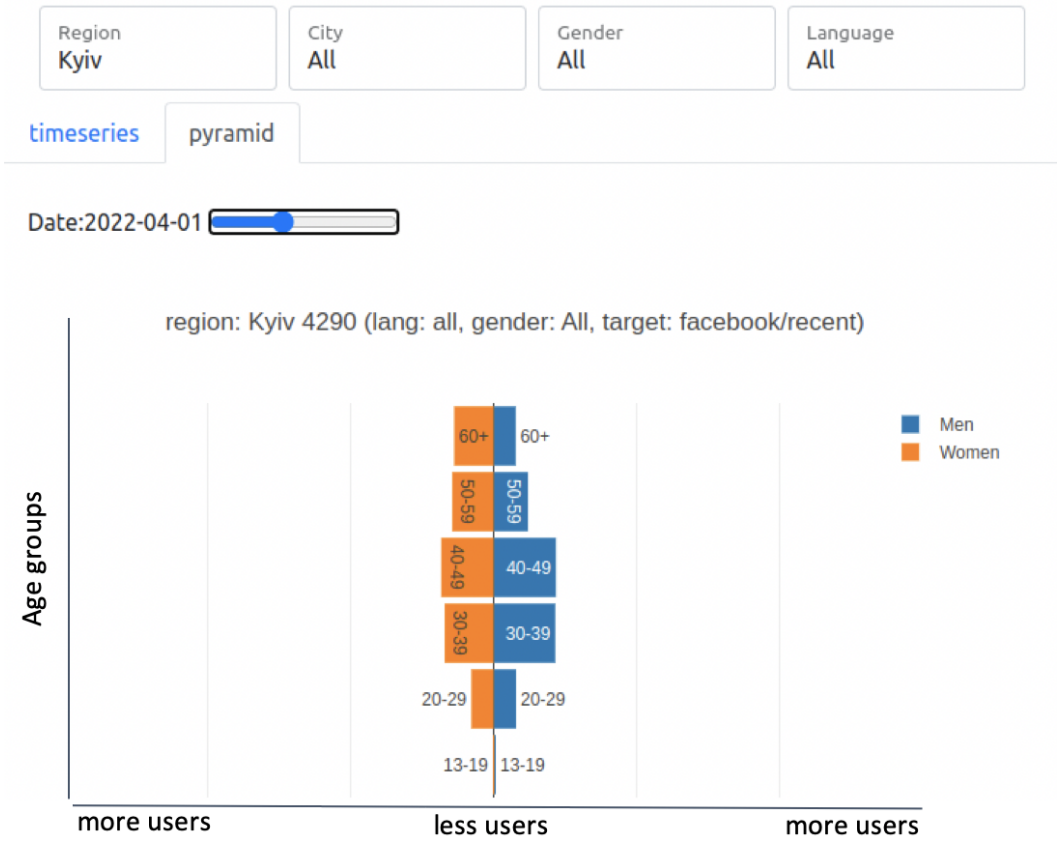
Ukraine


Facebook Marketing API
Request
- Location (e.g. Kyiv, Ukraine)
- Gender
- Age
- Language
Response
- Counts of daily active users
- Counts of monthly active users

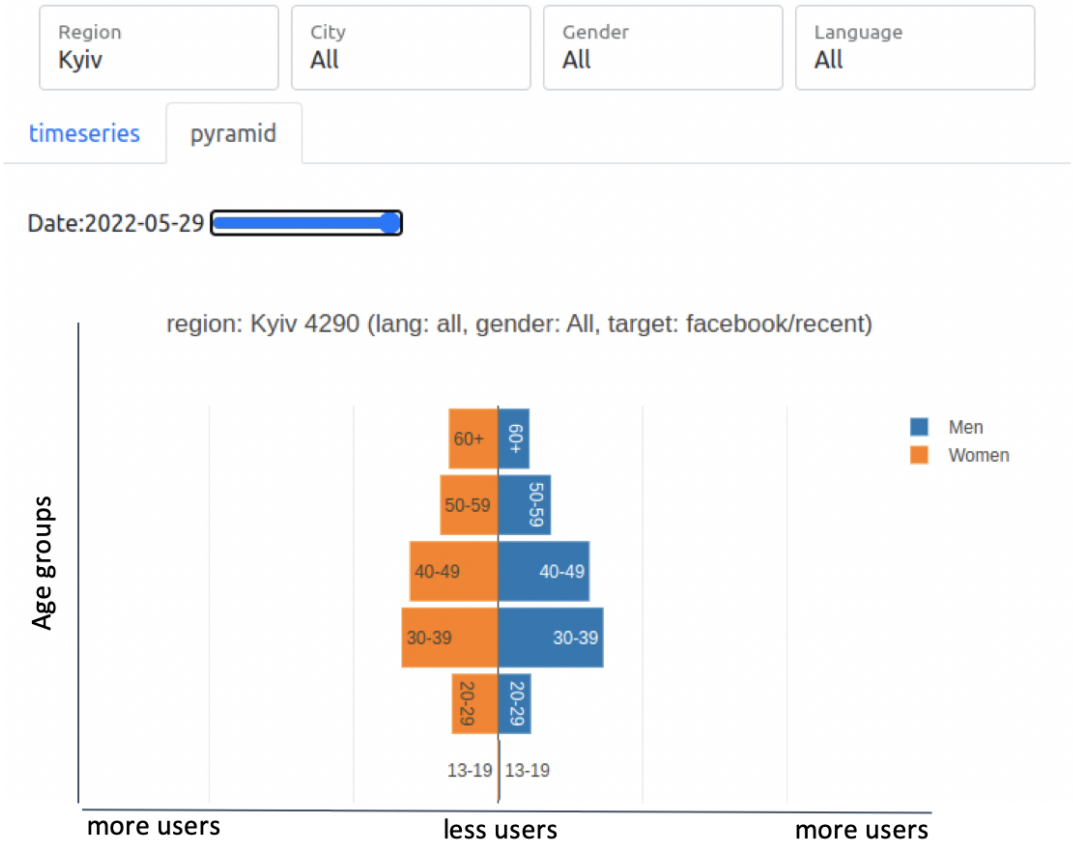
Ukraine


Region
Kyiv
City
All
Gender
All
Language
All
timeseries
pyramid
Date:2022-02-25
region: Kyiv 4290 (lang: all, gender: All, target: facebook/recent)
Men
Women
Age groups
60+
60+
50-59
50-59
40-49
40-49
30-39
30-39
20-29
20-29
13-19
13-19
more users
less users
more users

Ukraine


Region
Kyiv
City
All
Gender
All
Language
All
timeseries
pyramid
Date:2022-04-01
region: Kyiv 4290 (lang: all, gender: All, target: facebook/recent)
Men
Women
Age groups
60+
60+
50-59
50-59
40-49
40-49
30-39
30-39
20-29
20-29
13-19
13-19
more users
less users
more users

Ukraine

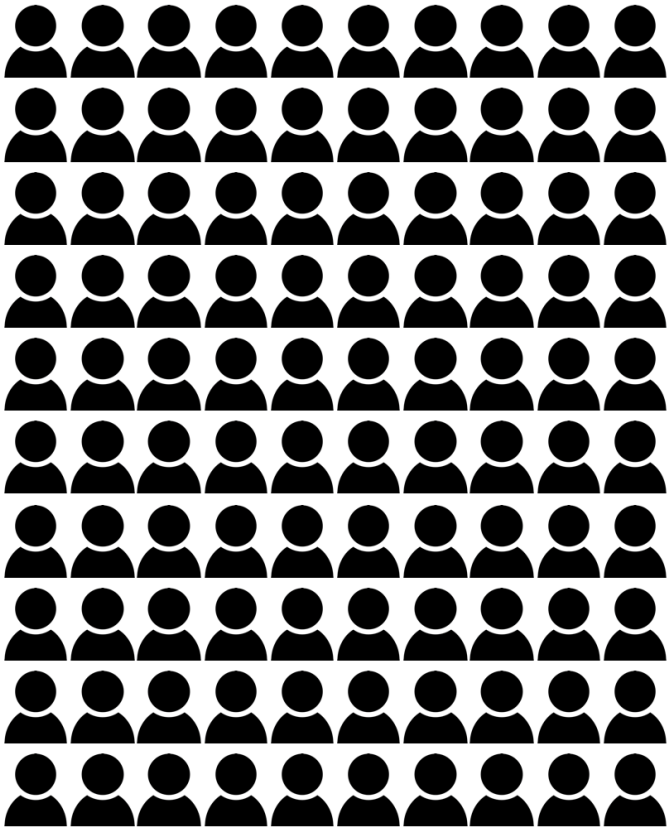

Region
Kyiv
City
All
Gender
All
Language
All
timeseries
pyramid
Date:2022-05-29
region: Kyiv 4290 (lang: all, gender: All, target: facebook/recent)
Men
Women
Age groups
60+
50-59
40-49
30-39
20-29
13-19
more users
less users
more users

From FB estimates to Population Counts

Population Estimation

30-34 year old women
in Kyiv before the conflict

Baseline population = 100



From FB estimates to Population Counts

Population Estimation

30-34 year old women
in Kyiv before the conflict

Baseline population = 100

Baseline Facebook users = 25

From FB estimates to Population Counts

Population Estimation

30-34 year old women
in Kyiv before the conflict

Baseline population = 100

Baseline Facebook users = 25

Baseline Facebook penetration = 25%

From FB estimates to Population Counts

Population Estimation

30-34 year old women
in Kyiv before the conflict

Baseline population = 100

Baseline Facebook users = 25

Baseline Facebook penetration = 25%

From FB estimates to Population Counts

Population Estimation

*How many 30-34 year old women are in Kyiv **today?***

Facebook users = 15

POPULATION = USERS / PENETRATION

From FB estimates to Population Counts

Population Estimation

*How many 30-34 year old women are in Kyiv **today?***

Facebook users = 15

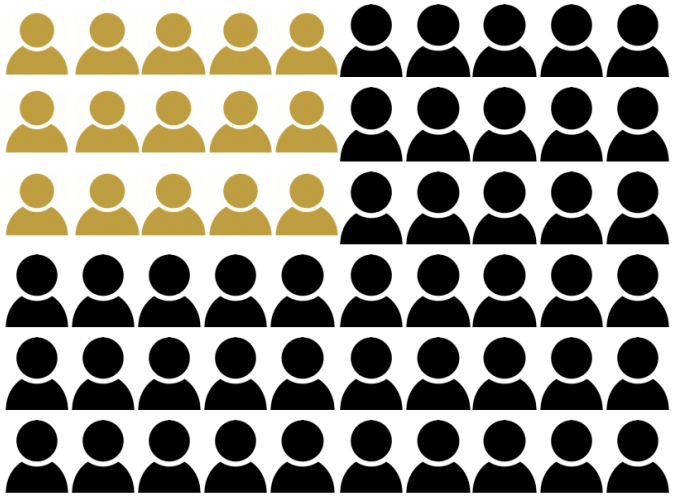
Baseline Facebook penetration = 25%

POPULATION = USERS / PENETRATION

From FB estimates to Population Counts

Population Estimation

*How many 30-34 year old women are in Kyiv **today?***



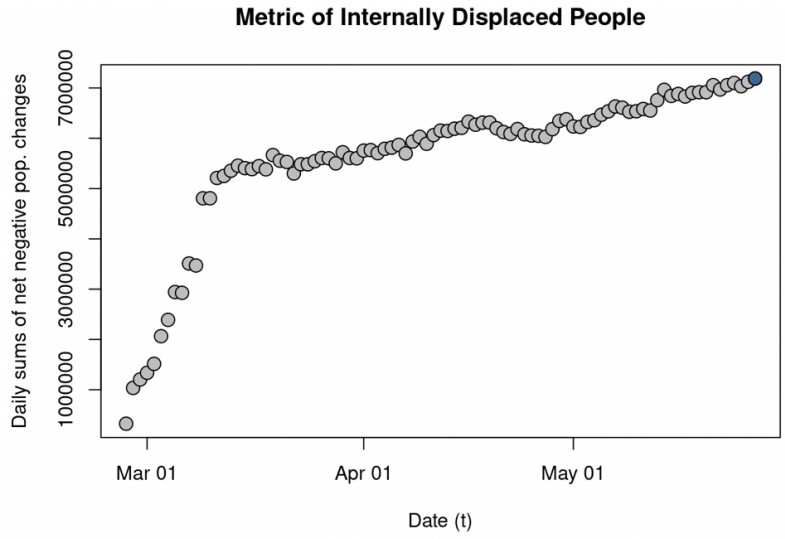
Facebook users = 15

Baseline Facebook penetration = 25%

Current population = 15 / 0.25 = 60

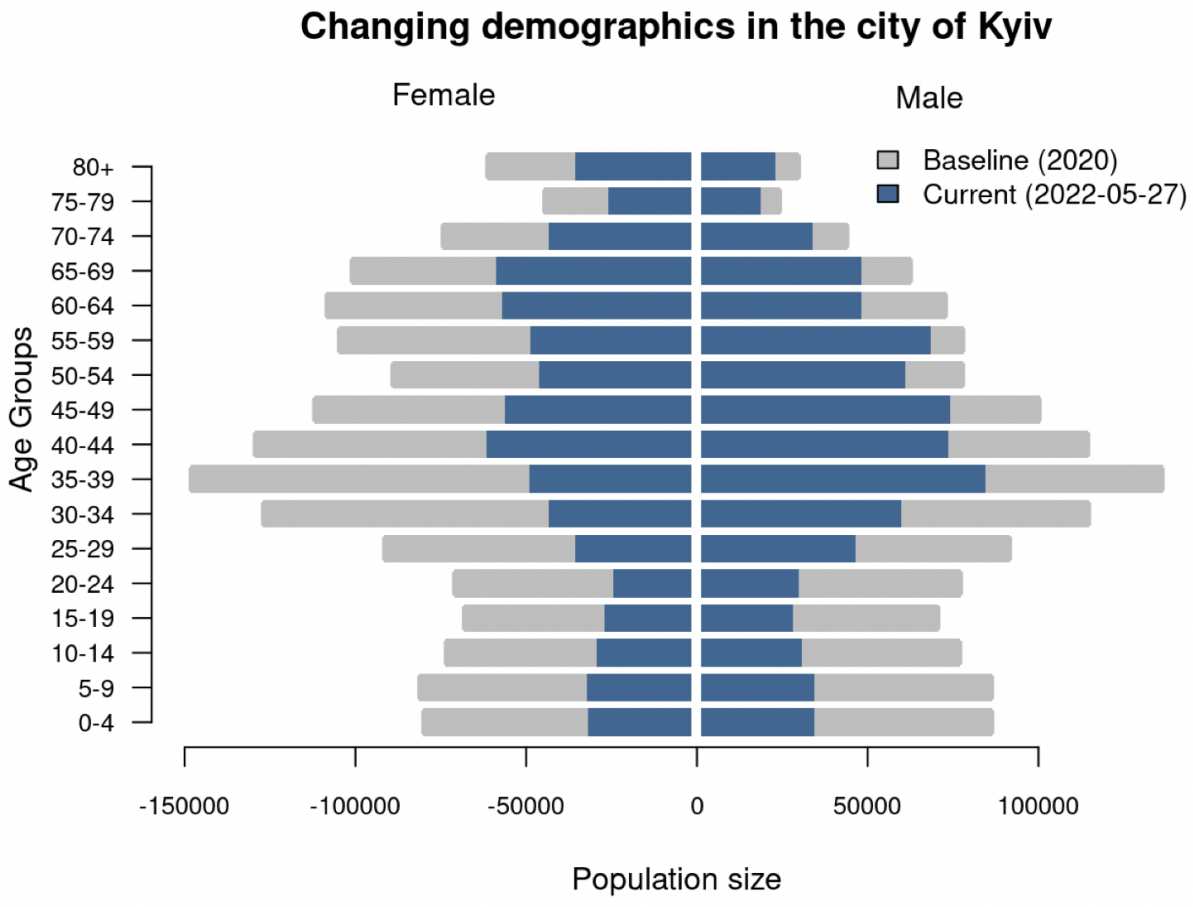
POPULATION = USERS / PENETRATION

Population Displacement Estimates



- As of May 27th: 7.2M fewer people estimated across all Oblasts where populations declined compared to baseline

Population Displacement Estimates



Summary: Social Media APIs

- Social media APIs serve as digital censuses that provide aggregated data on user populations.
- These data can provide valuable resolution for rapidly-evolving situations, or for monitoring social indicators.
- Yet, a significant shortcoming: no historic data available.
- Targeting categories change over time; their data generating process is opaque
- Algorithmic changes, version changes