

Data Science in International Economics Research

DSIER [/dɪ'zaiər/]

Julian Hinz & Irene Iodice

Bielefeld University

What this course is about

- Data science skills complementary to standard econometrics
 - Data cleaning and wrangling, visualization, databases, machine learning, etc.
 - Research in (broadly defined) International Economics shifts towards empirics
- We never had this course but wish did.

WHO WE ARE

WHO YOU ARE

NAME / PROGRAM / CODING BACKGROUND

Today's Roadmap

- Course overview
- Course logistics

COURSE OVERVIEW

WEEK 1

GETTING STARTED – COURSE OUTLOOK AND
GOOD RESEARCH PRACTICE

Week 1: Getting started

Good research practice, this afternoon:

- Reproducibility
- Versioning, Containerization and IDE
 - (Short) introduction to Git, Docker and VSCode
- Set up machine

EXPLORER

DSIER [DEV CONTAINER: DSIER DOCKER]
 > .devcontainer
 01-getting-started
 example.R
 .gitignore
 README.md

example.R

```
1  ###
2  # R example
3  # 220408
4  ###
5
6  if (!require(pacman)) install.packages("pacman"); library(pacman)
7  pacman::p_load(data.table)
8  pacman::p_load(ggplot2)
9
10 d = data.table(a = c(1:10),
11               b = c(2:11))
12 View(d)
13
14 ggplot(d) +
15   geom_point(aes(x = a, y = b))
16
```

PROBLEMS 1

OUTPUT

TERMINAL

PORTS 2

DEBUG CONSOLE

```
###
if (!require(pacman)) install.packages("pacman"); library(pacman)
r$> pacman::p_load(data.table)
r$> pacman::p_load(ggplot2)
r$> d = data.table(a = c(1:10),
                  b = c(2:11))
r$> View(d)
r$> ggplot(d) +
  geom_point(aes(x = a, y = b))
r$>
```

R Plot

bash

R Interactive

Dev Container: DSIER Docker main 0 0 1 2 R 4.1.3: 691

WEEK 2

THE TOOLKIT – R AND THE SHELL

Week 2: The toolkit — R and the shell

- R basics, 'data.table', 'tidyverse' and 'ggplot2'
- Basic helpful shell commands
- Make and more git

WEEK 3

WEB SCRAPING AND APIS

Week 3: Web scraping and APIs

Extracting data from a website

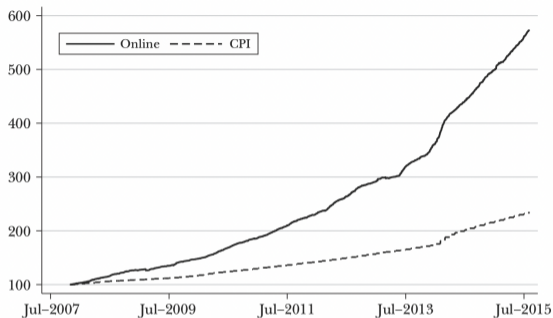
- Get the files for the website from a server
 - API
 - requests, web scraping
- Parse those files (CSS, HTML)

Week 3: Web scraping and APIs

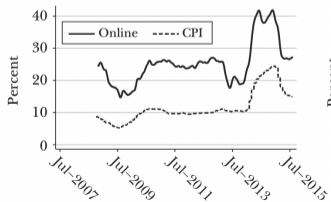
- “One Billion Prices Project”: Web-scraped prices for many stores and countries (Rigobon et al., 2016)
- Are online prices different than offline prices?
- Are reported CPIs still “true” today?

Argentina

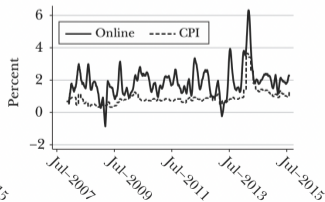
A: Price index



B: Annual inflation rate

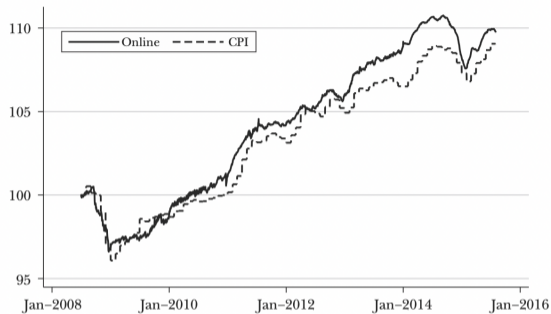


C: Monthly inflation rate

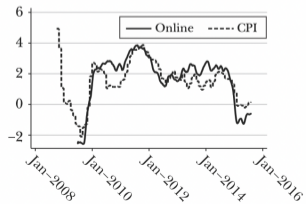


United States

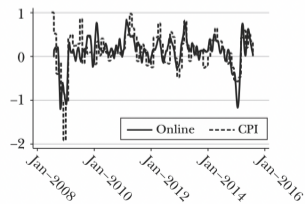
A: Price index



B: Annual inflation rate

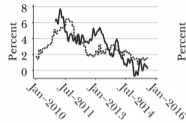


C: Monthly inflation rate

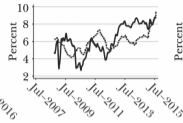


Online versus Consumer Price Index (CPI) Annual Inflation Rates

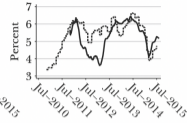
A: China



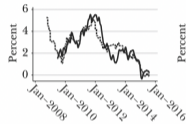
B: Brazil



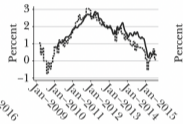
C: South Africa



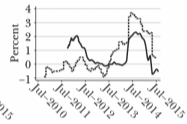
D: United Kingdom



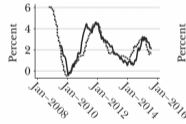
E: Germany



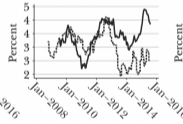
F: Japan



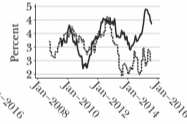
G: USA food



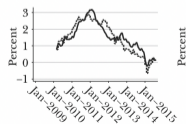
H: USA electronics



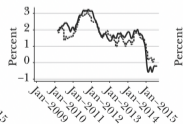
I: USA medical care



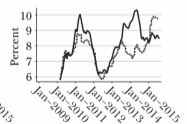
J: Eurozone



K: Developed markets



L: Emerging markets



— Online CPI

ikea.com

Planungsservice mit Termin

Aktuelle Informationen

Produktrückruf

IKEA

Was suchst du?

Hej! Jetzt anmelden.

Produkte

Räume

Angebote

Kundenservice

PLZ eingeben

Standort auswählen

Produkte

Aufbewahrung & Organisation

Regale & Allzweckregale

Bücherregale

BILLY Bücherregal

Bestseller









BILLY

Bücherregal, weiß, 80x28x202 cm

Preis inkl. MwSt.*

★★★★★ (274)

Wähle Farbe aus

Weiß

In den Warenkorb

Lieferverfügbarkeit prüfen

Verfügbarkeit für Click & Collect prüfen

Verfügbarkeit in deinem IKEA Einrichtungshaus prüfen

Montageservice & Installation Mehr erfahren

ikea.com

Was suchst du?

Produkte Räume Angebote Kundenservice

Produkte > Aufbewahrung & Organisation > Regale & Allzweckregale > Bücherregale > BILLY Bücherregal

Bestseller

BILLY
Bücherregal, weiß, 80x28x202 cm

Preis inkl. MwSt.*
★★★★★ (274)

45,00 €

PLZ eingeben Standort auswählen

Elements Console Sources Network Timelines Storage Graphics Layers Audit

div.pip-pip-price-package__m... span.pip-price span.pip-price_integer

```
<!DOCTYPE html>
<html lang="de-DE" dir="ltr" class="js-focus-visible js" data-js-focus-visible">
  <head>
    <body data-version="v111" data-shopping-available="true" data-list-available="true">
      <!-- 2022-04-15T14:15:09.714Z, Navigera fd28581c -->
      <a href="#hnf-content" class="hnf-skip-to-content hnf-btn hnf-btn--secondary" role="button"></a>
      <div class="hnf-messages"></div>
      <div class="hnf-header-hamburger hnf-page-container hnf-header-hamburger--sticky"></div>
      <header class="hnf-header hnf-header--store hnf-header--postalcode hnf-header--2nd-line"></header>
      <div id="hnf-header-filler" style=</div>
      <div class="hnf-page-container hnf-2nd-line"></div>
      <div class="hnf-location hnf-storepicker hnf-postalcodepicker hnf-page-container"></div>
      <aside class="hnf-menu hnf-menu--default hnf-menu--hidden" aria-hidden="true" data-title-more="Mehr"></aside>
      <div class="hnf-overlay"></div>
      <div id="hnf-content" tabindex="1"></div>
      <main class="main no-outline" role="main" id="content" tabindex="1">
        <div class="pip-page-container" data-cs-captures>
          <div class="pip-page-container_inner">
            <div class="pip-page-container_main">
              <div></div>
              <div class="pip-product_subgrid product-pip js-product-pip" data-product-id="00263850" data-product-no="00263850" data-product-type="ART" data-online-sellable="true" data-product-price="45" data-use-observe-product-details="true">
                <div class="pip-product_buy-module-container">
                  <div class="js-price-package pip-pip-price-package">
                    <div class="pip-pip-price-package_wrapper">
                      <div class="pip-pip-price-package_content-left"></div>
                      <div class="pip-pip-price-package_price-wrapper">
                        <div class="pip-pip-price-package_main-price">
                          <div class="pip-price">
                            <span class="pip-price_integer">45</span> = $0
                            <span class="pip-price_decimals"></span>
                            <span class="pip-price_currency-symbol pip-price_currency-symbol--trailing">€</span>
                          </div>
                        </div>
                      </div>
                    </div>
                  </div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </main>
    </body>
  </html>
```

Style Attribute {
range-pip-styles.12a3118b4efbda24189af.css:1:1...
W .pip-price_integer {
 line-height: 1.1875;
 font-size: 1.375rem;
 vertical-align: inherit;
}

Pseudo-Element :before
W *, main.fec5d203bod28aec231d.css:1:2102
*::before, *::after {
 box-sizing: inherit;
 outline: none;
}

Pseudo-Element :after
W *, main.fec5d203bod28aec231d.css:1:2102
*::before, *::after {
 box-sizing: inherit;
 outline: none;
}

Inherited From span.pip-price
range-pip-styles.12a3118b4efbda24189af.css:1:1...
+ Filter Classes

Box Model
position
margin
border
padding
x

Properties
box-sizing: border-box;
color: rgb(0, 0, 0);
display: inline;
font-family: "Noto IKEA", "Noto Sans", Roboto, "Open Sans", system-ui, sans-serif;
font-size: 22px;
font-stretch: normal;
font-style: normal;
font-variant-caps: normal;
font-weight: 700;
height: auto;

Auto — billy-buecherregal-weiss-00263850

WEEK 4

SOCIAL MEDIA DATA



Marvin | Runaways @ichmagdasnicht · 18 Apr 2017
Ich bin mit einer 2.0 in der Klausur einer der Schlechtesten. Ich studiere das Falsche, glaube ich langsam.

Translate from German

5

15



Timo Zander @tinkengil

Follow

Replying to @ichmagdasnicht

offenbar nicht Mathematik 📖

Translate from German

5:22 PM · 18 Apr 2017 from Kiel, Germany

1



Tweet your reply



Marvin | Runaways @ichmagdasnicht · 18 Apr 2017
Replying to @tinkengil
Wieso :D

1



Timo Zander @tinkengil · 18 Apr 2017
ich erinnere mich an eine Hororklausur mit 80% Durchfallquote. Bestnote war 1,7. Völlig absurd was da gefordert wurde

Translate from German

1

1



Marvin | Runaways @ichmagdasnicht · 18 Apr 2017
Gab's bei uns auch anfangs. Richtig Horror.

Translate from German

1





Timo Zander
@tinker11

Follow



Replying to @ichmagdasnicht

offenbar nicht Mathematik 🤖

Translate from German

5:22 PM · 18 Apr 2017 from **Kiel, Germany**

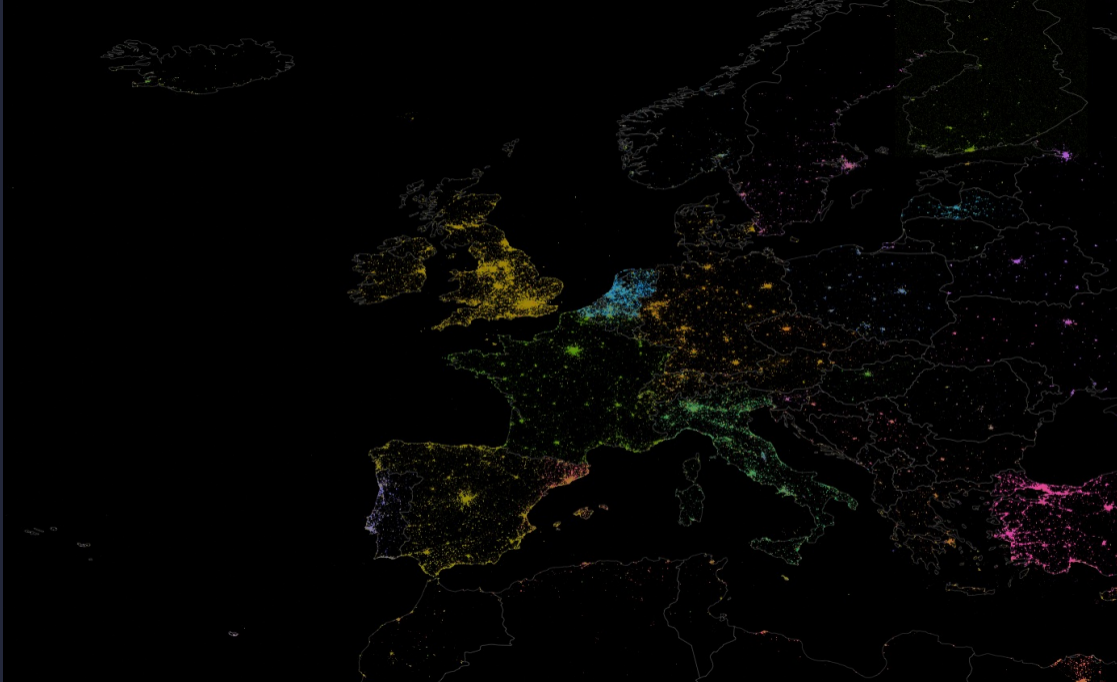


1



example_tweet.json


```
1 {
2   "created_at": "Tue Apr 18 15:22:19 +0000 2017",
3   "id": 854354410041991168,
4   "id_str": "854354410041991168",
5   "text": "@ichmagdasnicht offenbar nicht Mathematik 🍷",
6   "display_text_range": [
7     16,
8     43
9   ],
10  "source": "<a href=\"http://tapbots.com/tweetbot\" rel=\"nofollow\">Tweetbot for iOS</a>",
11  "truncated": false,
12  "in_reply_to_status_id": 854247992186073088,
13  "in_reply_to_status_id_str": "854247992186073088",
14  "in_reply_to_user_id": 2535411248,
15  "in_reply_to_user_id_str": "2535411248",
16  "in_reply_to_screen_name": "ichmagdasnicht",
17  "user": {
18    "id": 19030252,
19    "id_str": "19030252",
20    "name": "Timo Zander",
21    "screen_name": "tinkengil",
22    "location": "Kiel",
23    "url": "http://about.me/timozander",
24    "description": "PhD-Student | Podcastet bei playtogether-podcast.de | bloggt gelegentlich bei insulinaspekte.de und http://tinkengil.com | http://instagram.com/tinkengil",
25    "protected": false,
26    "verified": false,
27    "followers_count": 286,
28    "friends_count": 344,
29    "listed_count": 18,
30    "favourites_count": 1830,
31    "statuses_count": 12108,
32    "created_at": "Thu Jan 15 17:40:27 +0000 2009",
33    "utc_offset": 7200,
34    "time_zone": "Bern",
35    "geo_enabled": true,
36    "lang": "en",
37    "contributors_enabled": false,
38    "is_translator": false,
39    "profile_background_color": "EBEBEB",
40    "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/590786545/5vvyvdxrk528xhz91w06.jpeg",
41    "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/590786545/5vvyvdxrk528xhz91w06.jpeg",
42    "profile_background_tile": true,
43    "profile_link_color": "990000",
44    "profile_sidebar_border_color": "FFFFFF",
45    "profile_sidebar_fill_color": "F3F3F3",
46    "profile_text_color": "333333",
47    "profile_use_background_image": false,
48    "profile_image_url": "http://pbs.twimg.com/profile_images/549318880876048384/zaq6999H_normal.jpeg",
49    "profile_image_url_https": "https://pbs.twimg.com/profile_images/549318880876048384/zaq6999H_normal.jpeg",
50    "default_profile": false,
```

Week 4: Social media data

- Twitter, Facebook, Instagram, Spotify, ...
 - Hausmann & Hinz (2017), Bailey et al (2018), ...
- Human-recorded data: (Often) actual humans recording/being recorded what they see, think, listen to, where they are, who they are with, *en masse*

spotifycharts.com

 Charts

DOWNLOAD TO CSV

TOP 200














VIRAL 50

Filter by

GLOBAL

DAILY

04/18/2022

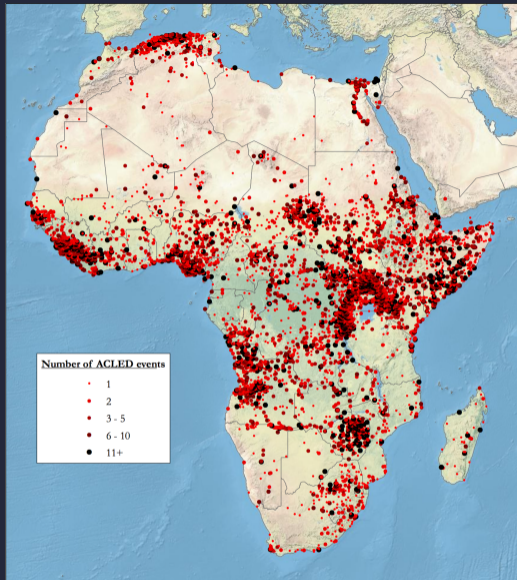
	TRACK	STREAMS ?
	1 – As It Was by Harry Styles	9,456,043
	2 – First Class by Jack Harlow	5,113,837
	3 – Heat Waves by Glass Animals	3,956,511
	4 – STAY (with Justin Bieber) by The Kid LAROI	3,131,056
	5 – Enemy (with JID) - from the series Arcane League of Legends by Imagine Dragons	2,841,206
	6 ▲ Bam Bam (feat. Ed Sheeran) by Camila Cabello	2,658,851
	7 ▼ Cold Heart - PNAU Remix by Elton John, Dua Lipa	2,615,083
	8 ▲ INDUSTRY BABY (feat. Jack Harlow) by Lil Nas X	2,549,615
	9 ▼ Envolver by Anitta	2,479,875
	10 – Una Noche en Medellín by Cris Mj	2,442,540
	11 ▲ Ghost by Justin Bieber	2,330,239
	12 ▼ abcdefu by GAYLE	2,325,809
	13 – MIDDLE OF THE NIGHT by Elley Duhé	2,289,566

WEEK 5

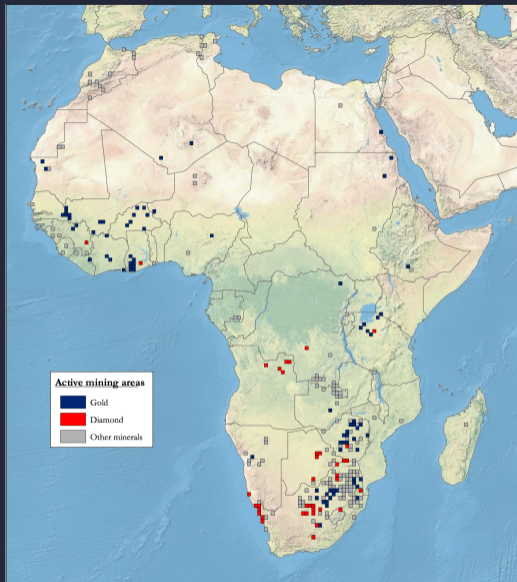
EVENT AND SENSOR DATA

Week 5: Event and sensor data

- **Event data** is any data that you want to measure about an event
- **Sensor data** is the output of a device that detects and responds to some type of input from the physical environment.
- Overview of the instruments to tackle **counterfactual analysis**



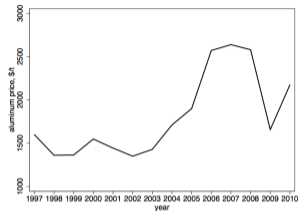
Source: Berman et al. (2017)



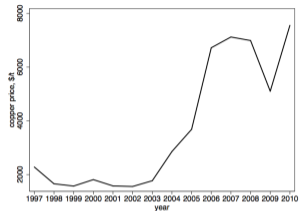
Source: Berman et al. (2017)

Figure A.6: Mineral prices (current USD)

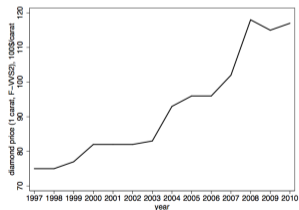
(a) Aluminum



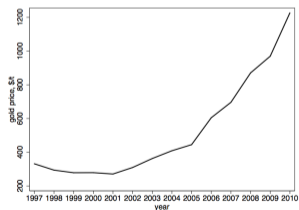
(b) Copper



(c) Diamond



(d) Gold



Source: Berman et al. (2017)

Week 5: Event and sensor data

Berman et al. (2017): “This Mine is Mine! How Minerals Fuel Conflicts in Africa”

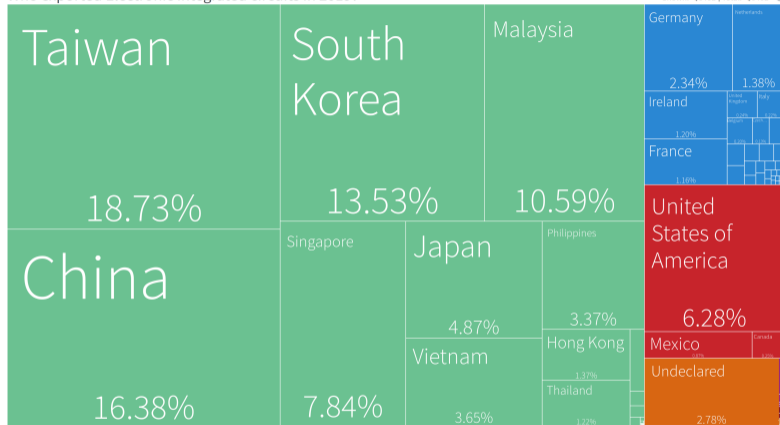
- Geolocalized data on conflict events in African countries between 1997–2010
- Geolocalized data on mining extraction of 14 minerals (Raw Material Data)
- Mining activity increases the incidence of conflicts at the local level
- then spreads violence across territory and time

WEEK 6

NETWORKS

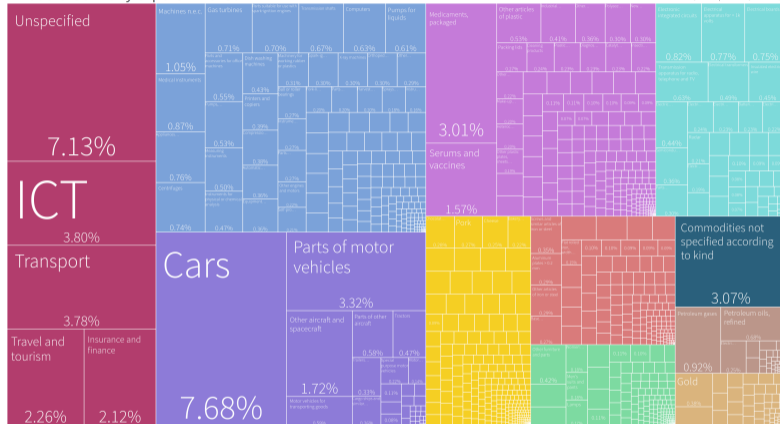
Who exported Electronic integrated circuits in 2019?

Shown: \$646B | Total: \$646B ⓘ



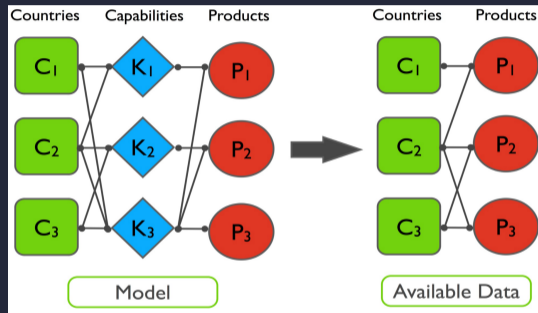
Browse more products here: <https://atlas.cid.harvard.edu/>

Shown: \$1.85T | Total: \$1.85T ⓘ



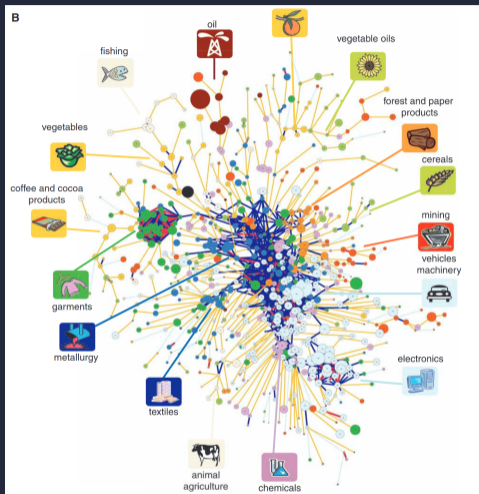
Browse more products here: <https://atlas.cid.harvard.edu/>

Week 6: Trade network



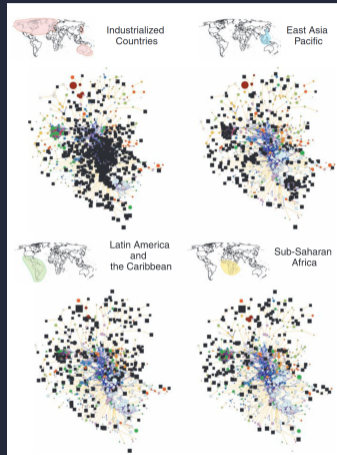
Source: Hidalgo et al. (2009)

Week 6: The Product Space of Trade



Source: Hidalgo et al. (2007)

Week 6: Countries in the Product Space



Source: Hidalgo et al. (2007)

Week 6: A glance at BACI

```
> library(vroom)
> vroom("~/data/BACI_HS12/BACI_HS12_Y2012_V202102.csv.gz") %>% head()
Rows: 8989697 Columns: 6
Delimiter: ","
chr (1): k # -> product HS6 digits
dbl (5): t, i, j, v, q # -> year, imp. ID, exp. ID, value, quantity
# A tibble: 6 × 6
   t      i      j k      v      q
<dbl> <dbl> <dbl> <chr> <dbl> <dbl>
1  2012      4      8 8 610469 2.73  0.044
2  2012      4      8 8 630900 0.11  0.082
3  2012      4     12 091099 1.18  1.2
4  2012      4     32 392630 1.30  0.123
5  2012      4     32 710310 0.096 0.004
```

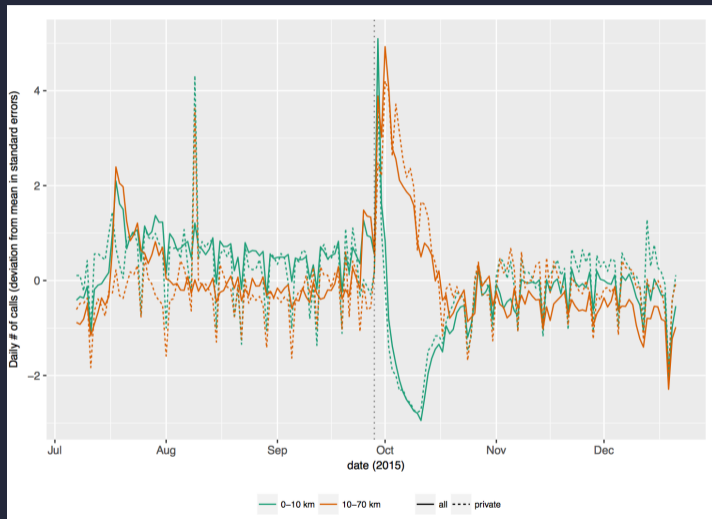
Week 6: Networks

1. How to handle trade data as a network
2. Which are useful metrics to capture country's performance from what they export?

WEEK 7

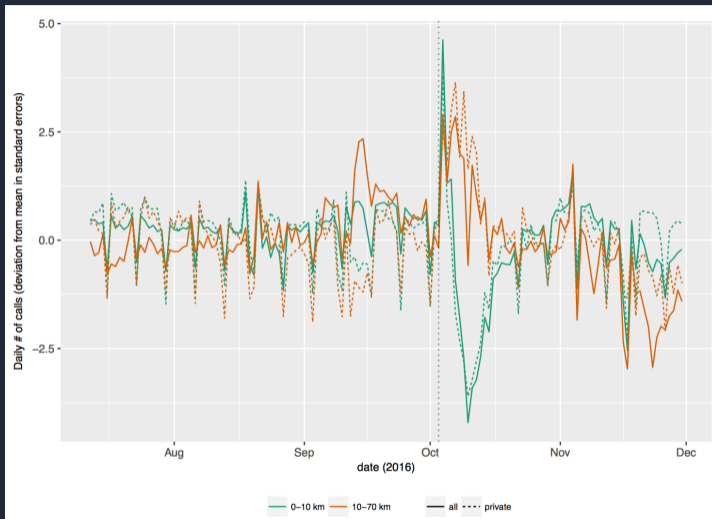
SPATIAL DATA

Week 7: Calls in Afghanistan



Source: Blumenstock et al. (2012)

Week 7: Calls in Afghanistan



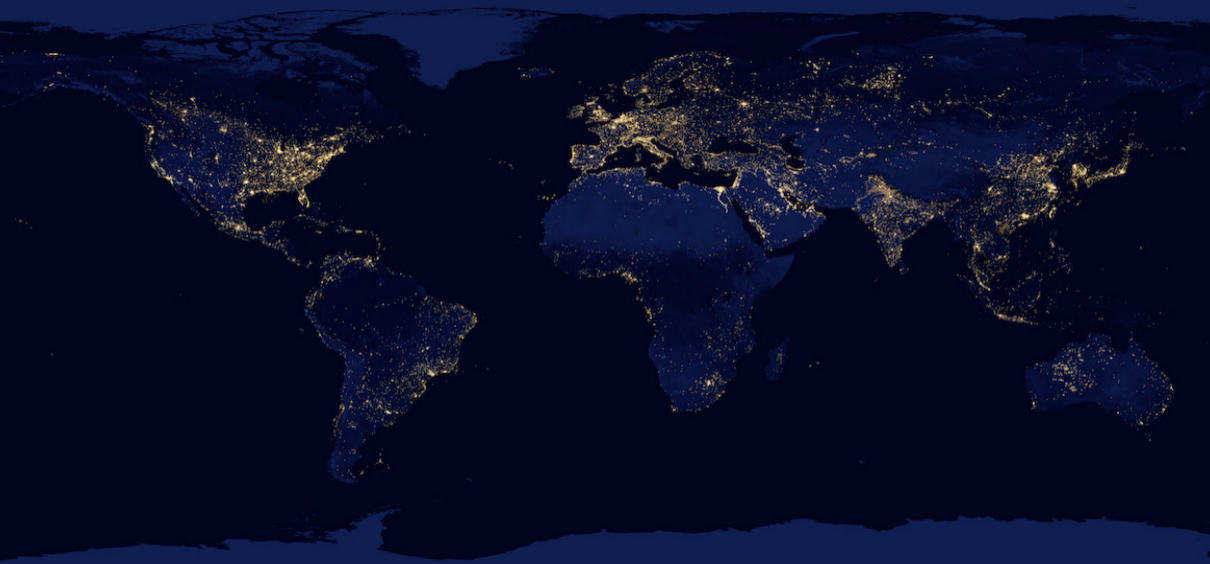
Source: Blumenstock et al. (2012)

Week 7: Spatial data

- Working definition: Units indexed time and space
- Call Detail Records
 - Blumentock et al. (2012): Firm activity after terror incidents
 - Ehrlich et al. (2018, 2020): Real life connections in urban environments
- many other spatial data: Taxi Rides, AIS data, ...

WEEK 8

SATELLITE IMAGERY





1992



1993

1994



1995



1996





1997



1998



1999

2000



2001



2002



2003



2004



2005



2006



2007



2008



2009



2010



2011



2012





1992

Week 8: Satellite imagery

- Satellite data: “Unbiased” recording of (human) activity / conditions / environment
- Nightlight data: Where do people live / consume / produce?
- Other data: Gas emissions, ship and airplane movements, ...

WEEK 9

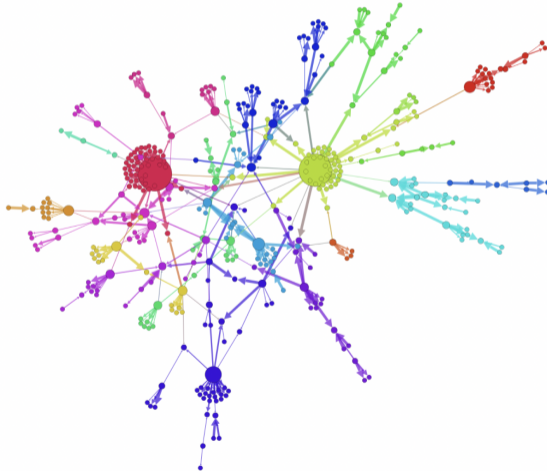
LARGE STRUCTURED DATA

Week 9: Conventional large data

- Trade not actually between countries, but firms in countries
- Bernard & Moxnes (2018): Vast majority of trade transactions includes at least one large firm with many trading partners
- Empirical regularities emphasizing firm heterogeneity among both buyers and suppliers

Week 9: Conventional large data

Figure 1: Firm-to-Firm Trade. U.S. importers and Norwegian exporters, HS 847990, 2006.



Week 9: Conventional large data

```
> pacman::p_load(data.table)
>
> data = fread("input/2019_Enero.csv.gz")
> head(data)
```

	FECH	ADUA	PAISGEN	PAISPRO	PAISCOM	DEPTODES	VIATRANS	BANDERA	REGIMEN
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<char>
1:	1901	87	169	17	169	8	1	43	C662
2:	1901	3	13	23	23	11	4	23	C100
3:	1901	48	17	23	23	11	1	169	C100
4:	1901	48	17	23	23	11	1	169	C100
5:	1901	3	17	23	23	11	4	23	C100
6:	1901	48	23	23	149	13	1	434	C130

	ACUERDO	PBK	PNK	CANU	CODA	NABAN	VAFODO	FLETE	VACID
	<int>	<char>	<char>	<char>	<char>	<i64>	<char>	<char>	<char>
1:	NA	16800	16800	16800	KG	506900000	10884,48	1350	12384,48
2:	NA	9,52	8,27	8,27	KG	3822009000	1718,5	229,93	1972,91
3:	NA	2,07	2,58	2	U	6241220000	87,04	2,05	04,75

Week 9: Conventional large data

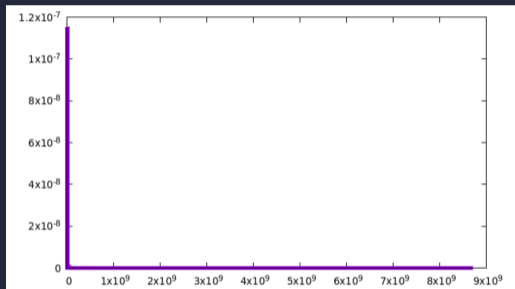
	VACIP	IMP1	OTDER	CLASE	CUIDAIMP	CUIDAEXP	ACTECON	CODADAD	VADUA
	<i64>	<i64>	<int>	<char>	<int>	<char>	<int>	<int>	<char>
1:	40559296	0	0	P	8001	HAMBURGO	0	87	12384,48
2:	6195312	1177000	0	P	11001	WIESBADEN	0	3	1972,91
3:	300482	66000	0	P	25126	WALDERSHOF	0	48	91,75
4:	690176	151000	0	P	25126	WALDERSHOF	0	48	210,74
5:	798770	152000	0	P	11001	HAMBURGO	0	3	254,67
6:	315664460	0	0	P	11001	TORONTO	0	48	100524
	VRAJUS	BASEIVA	OTROSP	OTROSBASE	TOTALIVAYO	SEGUROS	OTROSG	LUIN	
	<char>	<i64>	<int>	<int>	<i64>	<char>	<char>	<char>	
1:	0	40559296	0	0	0	150	0	Barranquilla	
2:	0	6195312	0	0	1177000	,26	24,22	Bogota	
3:	0	345482	0	0	141000	,89	0	Cartagena	
4:	0	794176	0	0	324000	2,06	0	Cartagena	
5:	0	798770	0	0	152000	,11	0	Bogota	
6:	0	315664460	0	0	0	125,63	24	Cartagena	

Week 9: Conventional large data

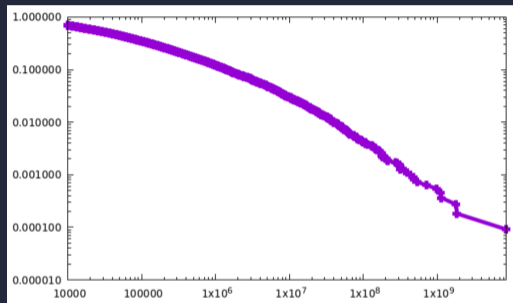
	CODLUIN	DEPIM	COPAEX	TIPOIM	PORARA	NIT	DIGV
	<char>	<int>	<int>	<int>	<char>	<i64>	<int>
1:	BAQ	8	17	1	0	901052265	6
2:	BOG	11	23	1	0	860002134	8
3:	CTG	25	23	1	15	900989716	4
4:	CTG	25	23	1	15	900989716	4
5:	BOG	11	23	1	0	800249678	5
6:	CTG	11	149	1	0	900023785	7

	RZIMPO	DEREL
	<char>	<int>
1:	AGROPECUARIA BARRAZA PACHECO SAS	0
2:	ABBOTT LABORATORIES DE COLOMBIA S.A.	0
3:	BICIJUAN COLOMBIA S.A.S.	45000
4:	BICIJUAN COLOMBIA S.A.S.	104000
5:	JUNGHEINRICH COLOMBIA SAS	0
6:	COLFOODS S A S	0

Colombian firms' export value



Arithmetic scale



Log-log scale

WEEK 10

TEXT AS DATA

Week 10: Sources of Text Data

Large availability of unstructured text-data from

- websites such as Twitter, Facebook, Google, and Wikipedia
- historical archives and administrative records

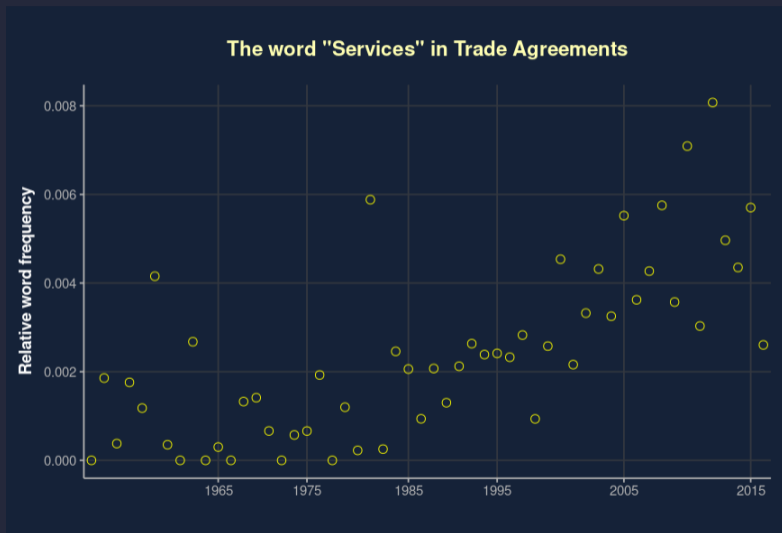
Week 10: Trade Agreements as Text Data

The treemap visualizes the structure of the ToTA corpus. The largest category is 'Administrative And Final Provisions' (yellow), followed by 'Institutions' (blue). Other significant categories include 'Dispute Settlement' (green), 'Rules Of Origin' (red), 'General Definitions' (orange), 'Intellectual Property' (light blue), 'Investment' (teal), 'Exceptions' (purple), 'Agriculture' (pink), 'Trade Remedies' (dark green), 'TBT' (brown), 'Transparency' (light yellow), 'Other' (light blue), 'E-Commerce' (light green), 'Socio-Cultural Provisions' (dark green), 'Objectives' (red), 'Trade In Goods' (yellow), 'Customs' (dark green), 'SPS' (dark blue), 'Financial Services' (orange), 'Financial Markets' (light orange), 'Labor' (light orange), 'Telecommunication' (light purple), 'Trade Liberalization Agency' (light purple), 'Multiple' (light purple), 'Trade Facilitation' (light purple), 'Energy' (light purple), 'Regulatory Cooperation' (light purple), 'Sustainable Development' (light purple), 'Textile' (light purple), 'Movement Of Natural Persons' (light purple), 'Environment' (light purple), 'Government Procurement' (light purple), 'Capital Movement' (light purple), 'Industrial Products' (light purple), 'Competition' (light purple), 'Inter-state Cooperation' (light purple), 'Trade In Services' (light purple), and 'General Provisions' (dark blue).

WorldCloud of the document of the Regional Comprehensive Economic Partnership



Week 10: Evolution of the content of Trade Agreements over time



Week 10: Overview of the topics

1. How to process and extract info from text data

2. Applications:

- 2.1 Measuring Policy Uncertainty from the News (Baker, Bloom et al. 2016)

- 2.2 Defining competitive clusters by text descriptions of products (Hoberg and Phillips 2016)

WEEK 11

OCR: DIGITIZED DATA

Week 11: OCR

- Information (data) recorded for a long time
 - Clay tables, papyrus scrolls, ...
- How to get that information at scale?
 - Optical character recognition
- Dell et al. (2020): “The Development Effects Of The Extractive Colonial Economy: The Dutch Cultivation System In Java”

11. Land.

Ferreslg C. Dessa's aan de onderneming dienstbaar, voor arbeid.

Namen der Breed's	Afkomst in pelen van de		Kultuur- directiepligte huizingen.	Eenmaalverlof door ieder deelen te verhoeden.	Groff kostend huizingen per huize.	Verwijzing aan de toelichtingen.
	coloniën- volken.	Indisch.				
<i>De transport</i>			402	62 3/4		
<i>Lade</i>		4 1/2	23	9	6	
<i>Handprangas</i>		1 1/2	1	7 1/2	13	
<i>Katigroetes</i>	3	6	50	4	7	
<i>De natural huizingen</i>	1	6 3/4	27	9	7	
<i>Spandjaren</i>	3	6	90	6 1/2	7	
<i>Bluen</i>	3	6 1/2	20	2 1/2	1	
<i>Samulaten</i>	1	6	11	15	7	
<i>Blanggrove</i>	1	6 1/2	8	3	6	
<i>De lade</i>	3	6 1/2	66	6 1/2	7	
<i>De boek de lade</i>	3	6	16	13	10	
<i>Blidge</i>	3	6 1/2	67	7 1/2	9	
<i>Wijze kullen</i>	1	6 1/2	60	9 1/2	7	
<i>Glas-smaling</i>	1	6 1/2	9	7 1/2	1	
<i>Kawee</i>	1	6 1/2	13	20	7	
<i>Jangpangrove</i>	1 1/2	6 1/2	8	1	1	
<i>De lade de lade</i>	1	6	10	2 1/2	6	
<i>De lade lade</i>	1 1/2	6	12	1 1/2	3	
<i>De lade kullen</i>	1 1/2	6 1/2	12	2	6	
<i>De lade kullen</i>	1 1/2	6 1/2	22	6	6	
<i>De lade lade</i>	1	6	3	3	6	
<i>De lade lade</i>	1	6 1/2	11	2 1/2	7	
<i>De lade lade</i>	3	6 1/2	8	1	3	
<i>De lade lade</i>	3	6 1/2	10	1 1/2	7	
<i>De lade lade</i>	1	6 1/2	20	2 1/2	7	
<i>De lade lade</i>	1	6 1/2	26	2 1/2	7	
<i>De lade lade</i>	3	6 1/2	9	1	6	
<i>Transporten</i>	1		983	18 1/2		

H. Land.

Fervolg C. Dossa's aan de onderneming dienstbaar, voor arbeid.

Naam der Breed's.	Afstand in palen van de		Kultu- resdigthijde bezigrazen.	Baanverreist door ieder-don- te onthouder.	Gras besont kingslaam per bouw.	Verrijping van de teeltwijzen.
	submer- rieden.	diep.				
De transport	1	1	25	198 3/4		
Umschakelen	3/4	9 3/4	10	2	5	
Overvoering	1	2 1/2	11	2 1/2	5	
Wetervoor	3/4	4	9	1 1/2	6	
Opvoerbakering	3/4	2 1/2	2 1/2	2 1/2	7	
Bakering	3/4	2 1/2	2 1/2	2 1/2	6	
Ammeringen	3/4	2	1 1/2	2	6	
Stadskakering	1	2	2 1/2	6 1/2	7	
Stadskakering	1	2 1/2	2 1/2	5	6	
Stadskakering	3/4	9 3/4	2 1/2	5	6	
Stadskakering	1	2 1/2	1 1/2	2	8	
Stadskakering	1 1/2	6	1 1/2	2 1/2	7	
Stadskakering	2	6	9	1	9	
Stadskakering	2	2 1/2	11	1 1/2	7	
Stadskakering	1 1/2	2	2 1/2	5	7	
Stadskakering	1 1/2	2 1/2	2 1/2	2	8	
Stadskakering	3/4	2 1/2	1 1/2	2	7	
Stadskakering	2	2 1/2	2 1/2	5	7	
Stadskakering	3/4	9 3/4	2 1/2	2 1/2	8	
Stadskakering	1	2	9 1/2	6 1/2	7	
Stadskakering	2 1/2	1 1/2	1 1/2	2	7	
Stadskakering	3/4	1	1 1/2	1 1/2	8	
Stadskakering	2 1/2	1 1/2	1 1/2	1 1/2	7	
Stadskakering	2 1/2	1 1/2	8	1 1/2	7	
Stadskakering	2	1	2 1/2	4	6	
Stadskakering	1 1/2	3/4	8	2 1/2	6	
Stadskakering	3/4	3/4	11	2	6	
Stadskakering	1	1	19 1/2	2 1/2 3/4		

Figure 1: Counterfactual Factories Example

(a) Real Factory

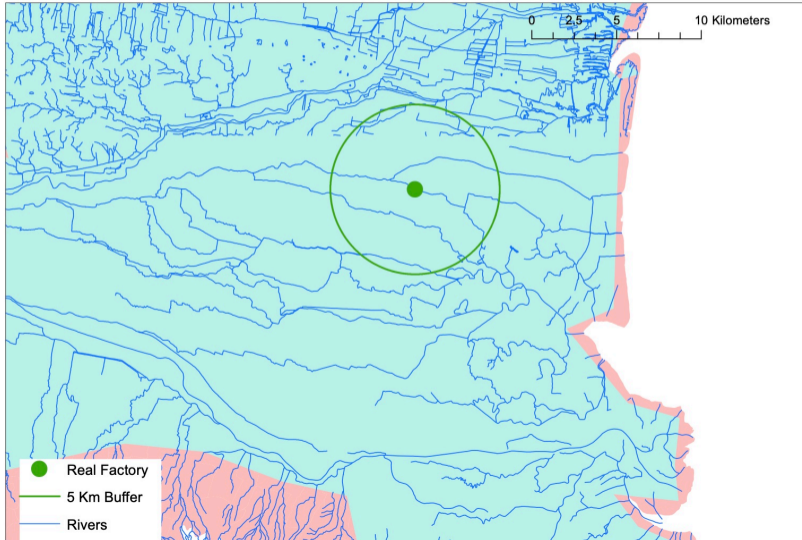
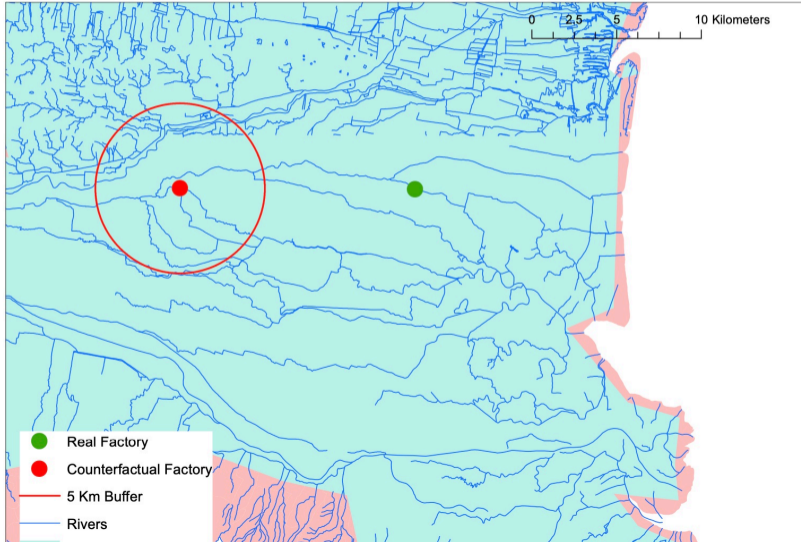


Figure 1: Counterfactual Factories Example

(b) Counterfactual Factory Suitability





PubLayNet



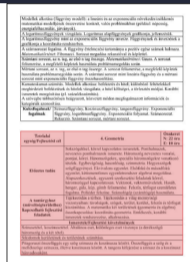
HJDataset



PRImA



Newspaper Navigator



TableBank

COURSE LOGISTICS

Schedule

- Morning session: Introduction to topic and relevant paper(s)
- Afternoon session: Hands-on session with the data
 - Live-coding
- In-person and via Zoom: <https://jhi.nz/dsier>

Coursework

Course project

- Choose a data (unconventional) dataset and either
 - Open access AWS data
- Replicate a paper *or*
- Test own hypothesis
- Final product: Report, website, visualizations... and presentation
- Twist: Git repository for all to see

Coursework

Course project

- Choose a data (unconventional) dataset and either
 - Open access AWS data
- Replicate a paper *or*
- Test own hypothesis
- Final product: Report, website, visualizations... and presentation
- Twist: Git repository for all to see

Further resources

- Website: dsier.julianhinz.com
- Slack: <https://dsier23.slack.com>
- Lots of references on the website

QUESTIONS?

SUGGESTIONS?

Data Science in International Economics Research

DSIER [/dɪ'zaiər/]

Julian Hinz & Irene Iodice

Bielefeld University