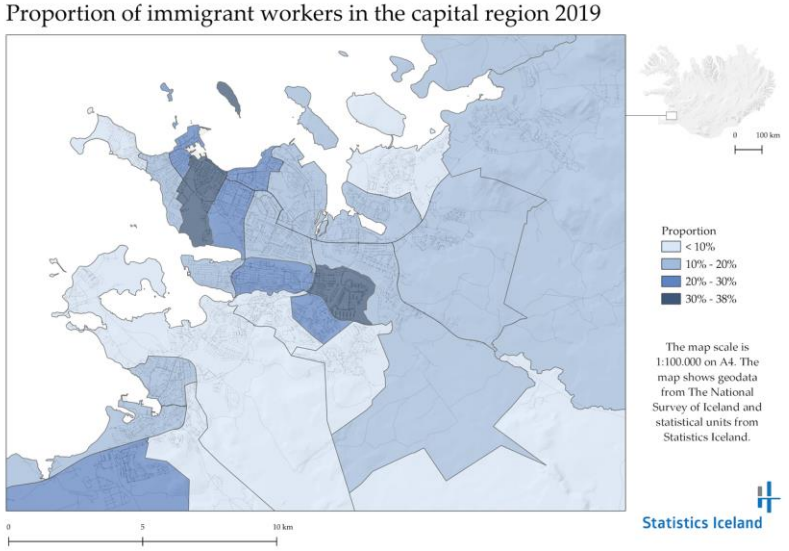
Using small statistical units to abstract commuting networks in Reykjavík, Iceland

Gísli Pálsson
Violeta Calian
Ómar Harðarson

Statistics Iceland

Employment statistics: leveraging a powerful resource

- Statistics Iceland publishes monthly employment statistics based on tax registers, with a lag of c. 2 months.
- These publications combine tax register data and demographic data to show several dimensions of Iceland's labour force.
- Geospatially, the data are divided into standard cadastral units as well as a recently completed set of statistical units.
- These statistical units follow current cadastral boundaries but divide more densely populated places to create areas of c. 1500 individuals.
- More on these later.



Experimental objective: A model of commuting routes

- As tax registers contain address information for employers, it is possible to model travel routes between an employee's home and their workplace.
- This is an aspect of employment register data that has currently not been available in Iceland.

The initial model methodology

Data

- Tax register data
- Demographic data
- Business register data
- OpenStreetMap data

Technology

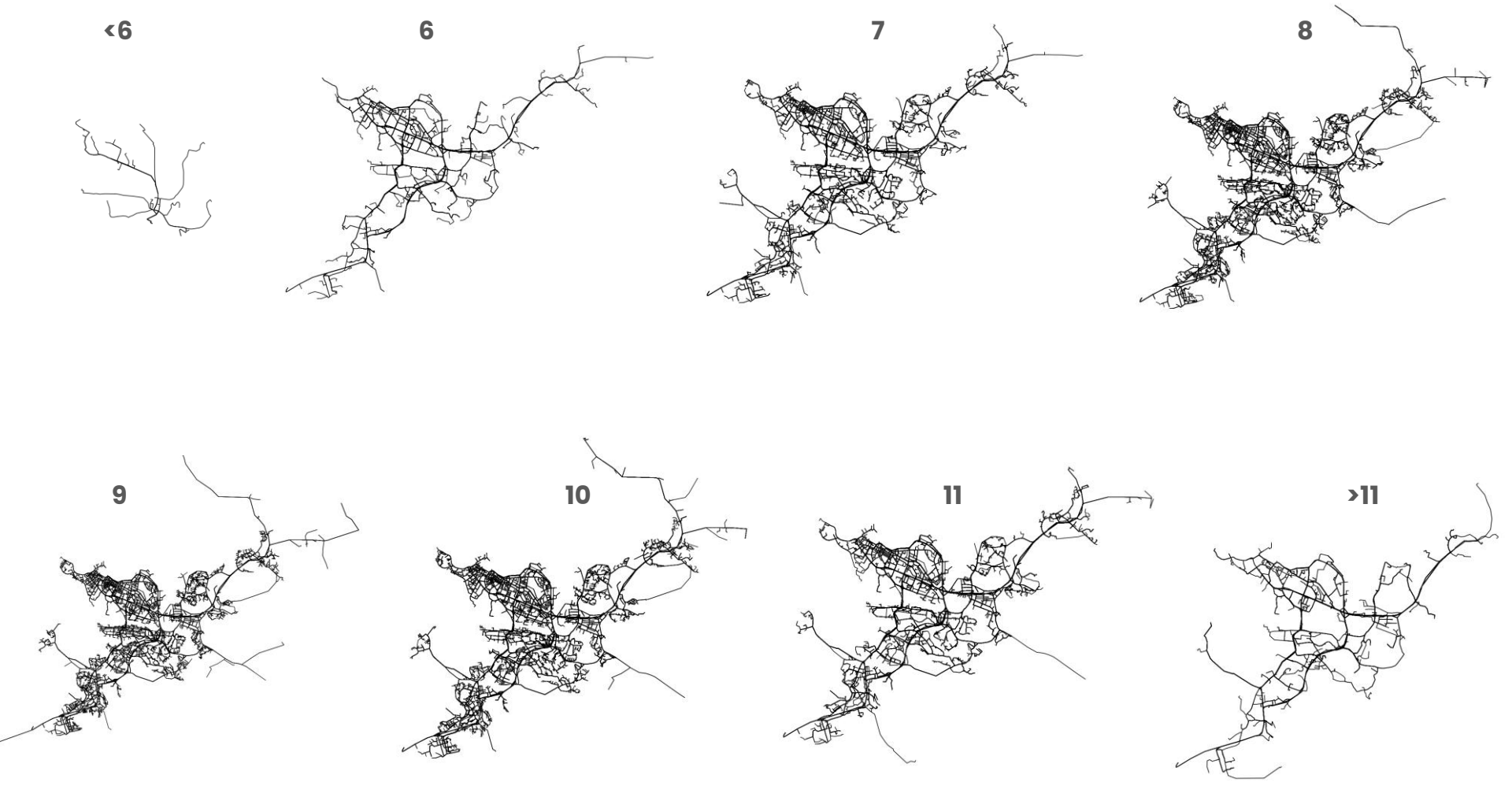
- PostgreSQL with PgRouting
- Python

```
# define login function
def login(password,user='gisli',dbase='hax_mapping',schema='public'):
    connection = psycopg2.connect(host=pg_host, port=pg_port, dbname=dbase, user=user, password=password,options=f"-c sea
    connection.autocommit = True
    engine = create_engine(f'postgresql://{user}:{password}@{pg_host}/{dbase}')
    return connection,engine
# connect to database
connection, engine = login(password)
# define source and target node table
nodes = pd.read_sql('select routing_smasv_source::integer,routing_smasv_target::integer from manntal_employment',engine)
nodes_clean = nodes.dropna()
# specify route building function
def build_route(source,target):
    cursor = connection.cursor()
    cursor.execute(f"""
with dijkstra as (
SELECT * FROM pgr_dijkstra(
    'SELECT id, source, target, cost, reverse_cost FROM iceland_2po_4pgr',
    {source}::integer,{target}::integer,
    FALSE)),
cte as (
select a.node, a.agg_cost, b.the_geom from dijkstra a inner join iceland_2po_4pgr_vertices_pgr b on a.node = b.id
)

insert into hax_smasv_network
select {source} as hax_source,{target} as hax_target, max(agg_cost), st_makeline(the_geom) from cte
""")
# run route building function
for i,j in nodes_clean.iterrows():
    build_route(j[0],j[1])
```

Methodology

- Lateral join between home and work coordinates and nearest node in the OSM transport network.
- Dijkstra's algorithm (pgr_dijkstra) for the shortest path.
- Assumes automotive travel.
- Assumes employees travel to their employer's headquarter office.
    - We have modelled this to direct employees to the nearest office up to a maximum limit, but are waiting to implement this as the underlying data will hopefully improve in the near future.
- Currently undergoing an estimate of the most common start and end times per economic activity class (based on NACE Rev. 2); current proxy uses a gaussian distribution for the start of a workday with a mean of 9 and standard deviation of 1 (this is just a rough estimate).

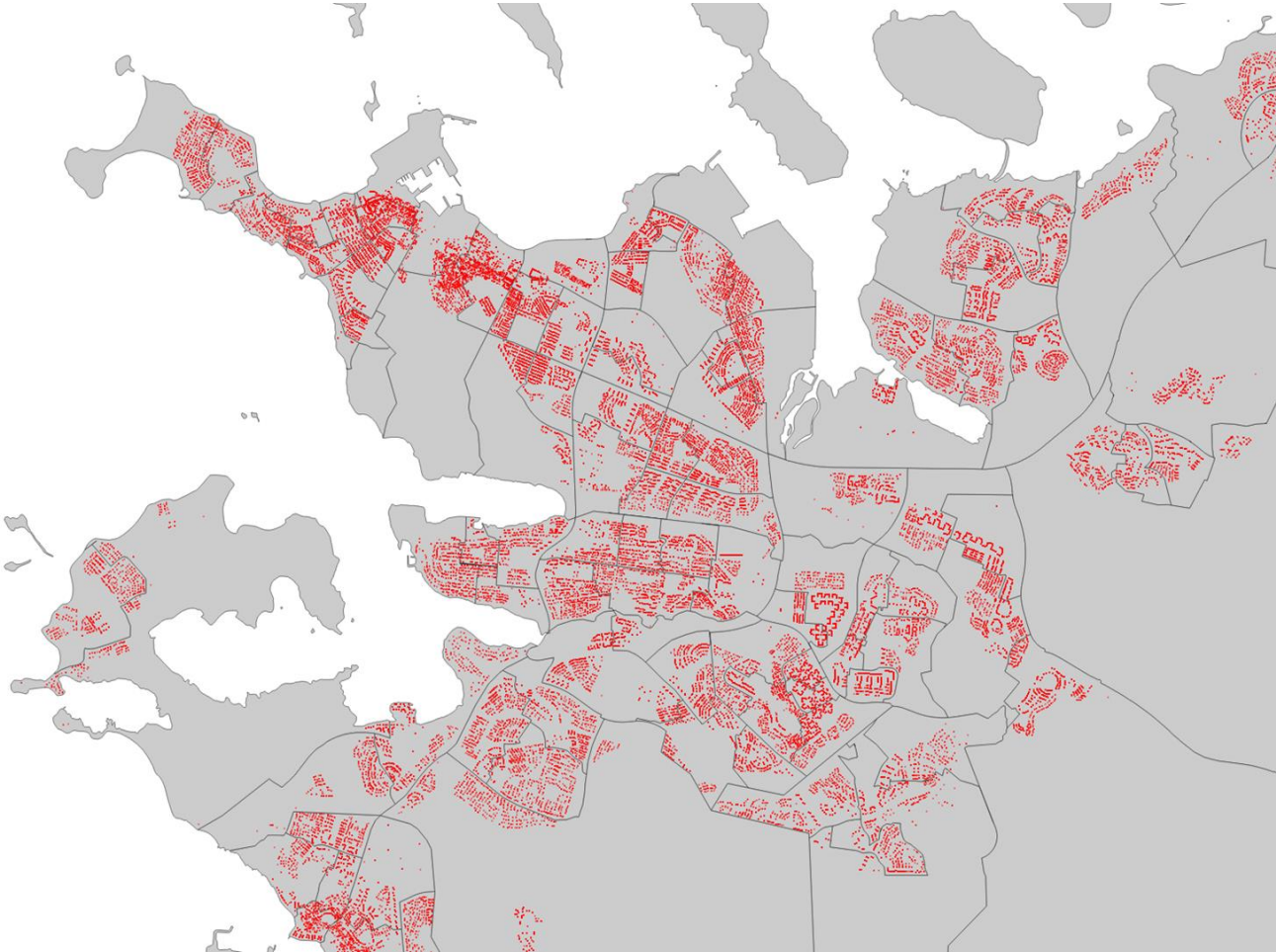

<6
6
7
8
9
10
11
>11

Personal data and the first model

- While this iteration of the commuter model has a lot of value for internal and sanctioned use, Statistics Iceland is unable to provide the model to the general public as it reveals too much information about persons in the workforce; namely, their home.

Response: Using small statistical units

- To provide a less individuated model, we grouped both home and work addresses by small statistical units
- These follow established cadastral units and divide the population into c. 1500 individuals per polygon.

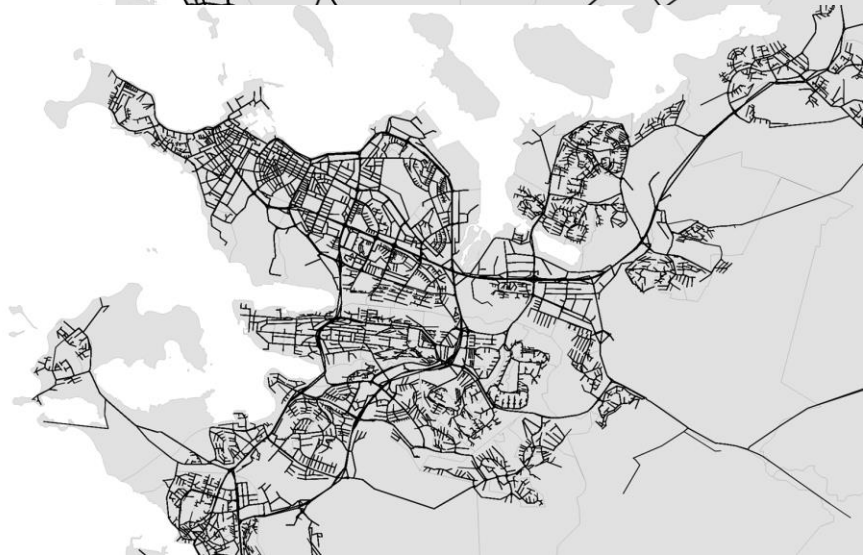


Response: Using small statistical units

- This more generalized model counts the same number of network edges but assumes centroid-to-centroid travel (using the PostGIS st_centroid function), still following Reykjavík's road network.
- Provides a general, publicly accessible version of the restricted model
- Also a useful generalization for understanding workforce mobility (and stress points)

Generalized model

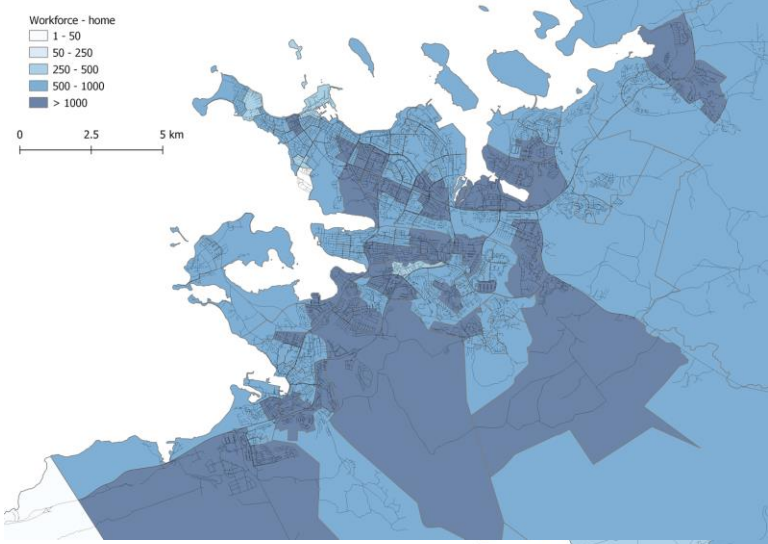
Individuated model



Generalized workforce flow

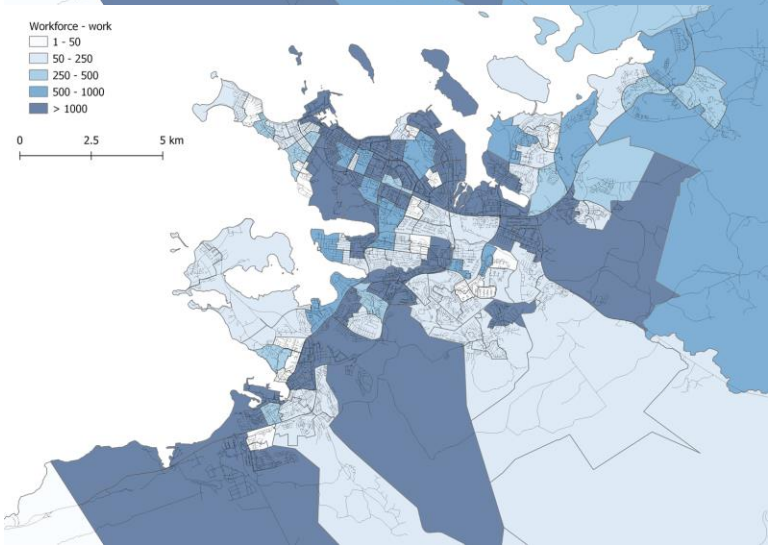
- This level of abstraction can show aspects that are more easily transferred to policy, for instance showing the displacement (and concentration) of the workforce into specific statistical units with high economic activity.

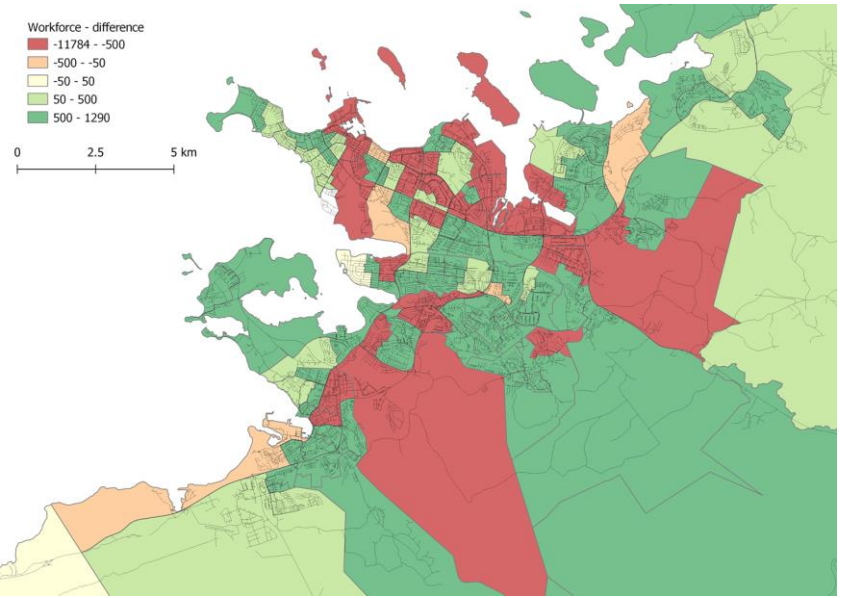
Workforce density:



Home

Work

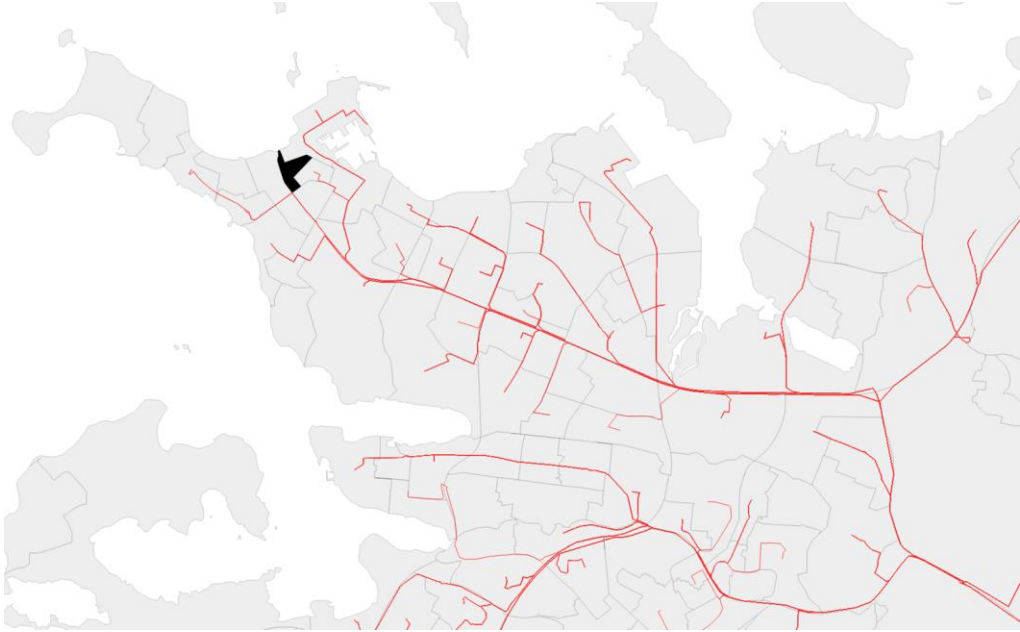




Workforce flow

Conclusions: A topology of the 2021 Census

- This project is a side effect of the work already underway to synthesize register-based data for the upcoming 2021 Census.
- While some improvements need to be made, this is a relatively low hanging fruit given the work already done.
- This project has been partially supported by the grant number 831732 — 28-IS-Merging (on Higher resolution Icelandic Statistical Geography Standard for register based census statistics).



Geospatial view

Sankey diagram