

Spatial Data Science in Python

micro-credentials course



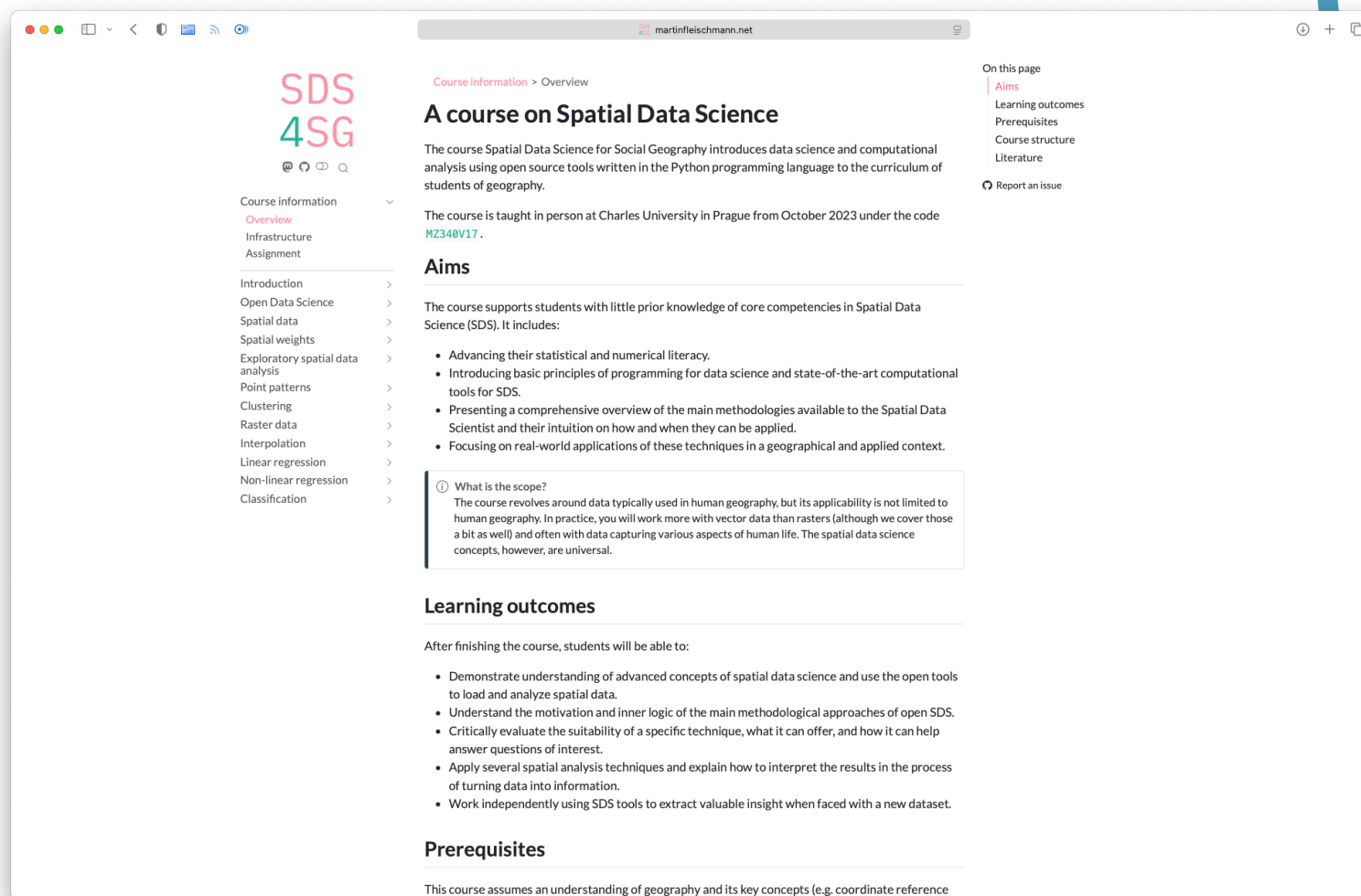
Výchozí situace



Doporučené volitelné předměty

Kód	Název	Rozsah	Kr.	Sem.
<i>MZ340V17</i>	<i>Spatial Data Science v sociální geografii</i>	<i>1/2 Zk</i>	<i>4</i>	<i>Z</i>





<https://martinfleischmann.net/sds/>





13 lekcí

1 lekce = 3 výukové hodiny

Primárně zaměřeno na magisterské studenty geografie



Motivace





Poptávka

Minimální investice

Příjem



Spatial Data Science in Python





10 ze 13 lekcí

Výuka plně angličtině

Online forma

Poplatek 300 Euro





Intenzivní forma výuky

2 týdny, jedna lekce denně, s jedním týdnem volna mezi

Zakončení stejné jako standardní kurz





Rozšíření statistické a numerické gramotnosti.

Seznámení s principy programování pro data science a nejmodernějšími výpočetními nástroji pro SDS.

Představení uceleného přehledu hlavních metodik, které má geograf v oblasti prostorových dat k dispozici, a vybudování intuice, jak a kdy je lze použít.

Zaměření na reálné aplikace těchto technik v kontextu sociální geografie.



Každá lekce rozdělena do tří částí - *Concepts, Hands-on a Exercise*



Předmět je zakončen seminární prací ve formě **výpočetní eseje**. Výpočetní esej je esej, jejíž narativ je podpořen kódem a jeho výsledky, které jsou součástí eseje



20 studentů, primárně ze zemí EU





martinfleischmann.net/sds/micro/



Course information >

Introduction >

Open Data Science >

Spatial data >

Spatial data old and new >

Bring in geography >

Learning GeoPandas >

Spatial weights >

Exploratory spatial data analysis >

Point patterns >

Clustering >

Raster data >

Interpolation >

Regression >

Making maps

Maps in GeoPandas are of two kinds. Static images and interactive maps based on [leaflet.js](#).

Static maps

GeoPandas can also plot maps to check how the geometries appear in space. Call `GeoDataFrame.plot()` to plot the active geometry. To colour code by another column, pass in that column as the first argument. In the example below, you can plot the active geometry column and colour code by the `"distance_to_arc"` column. You may also want to show a legend (`legend=True`).

```
_= buildings.plot("distance_to_arc", legend=True)
```

Map of distances to the Arc de Triomphe

The map is created using the `matplotlib` package. It's the same that was used under the hood for all the plots you have done before. You can use it directly to save the resulting plot to a PNG file.

```
import matplotlib.pyplot as plt
```

1 Import the `pyplot` module of the `matplotlib` package as the `plt` alias.

If you now create the plot and use the `plt.savefig()` function in the same cell, a PNG will appear on your disk.

```
buildings.plot("distance_to_arc", legend=True)
plt.savefig("distance_to_arc.png", dpi=150)
```

Other resources for static plotting

On this page

- Datasets
- Reading files
- Writing files
- Geometries
- Projections
- Simple accessors and methods
- Making maps
- Static maps**
- Interactive maps
- Spatial predicates
- Spatial join
- Overlay

Report an issue

Other Formats

Jupyter



The background features several thick, flowing blue lines in various shades, creating a sense of movement. In the bottom-left corner, there is a small white rectangular box with a thin grey border. Inside this box, at the top, are three icons: a circular arrow (refresh), a plus sign, and a document icon. The rest of the box is empty.

Course information >

Introduction >

Open Data Science >

Spatial data >

Spatial weights >

Exploratory spatial data analysis >

Point patterns >

Clustering >

Raster data >

Interpolation >

Regression >

Relationship and prediction >

Regression and geography

Fit that model

Moran Plot

That looks like a pretty strong relationship. Use the local version of Moran's statistic to find out the clusters.

```
lisa = esda.Moran_Local(elections_data['residual'], contiguity_r)
```

1 Use `Moran_Local` function from `esda`.

Let's visualise the results.

```
lisa.explore(elections_data, prefer_canvas=True, tiles="CartoDB Positron")
```

LISA clusters

The outcome of LISA shows large clusters of both overpredicted (High-High) and underpredicted (Low-Low) areas. The underpredicted are mostly in central Bohemia around Prague and in the mountains near the borders, where the ski resorts are. Putting aside the central areas for a bit, the explanation of underprediction in mountains is relatively straightforward. The education data are linked to the residents of each municipality. The people who voted in a municipality do not necessarily need to match with residents. It is known that more affluent population groups, who are more likely to go to a ski resort, voted overwhelmingly for Pavel. And since the elections were in winter, a lot of them likely voted in ski resorts, affecting the outcome of the model.

The overpredicted areas, on the other hand, are known for higher levels of deprivation, which may have played a role in the results. What is clear, is that geography plays a huge role in the modelling of the elections.

On this page

- Data
- Non-spatial linear regression
- OLS model
- Spatial exploration of the model (hidden structures)
- Spatial heterogeneity
- Geographically weighted regression

Report an issue

Other Formats

Jupyter

Course information >

Introduction >

Open Data Science >

Spatial data >

Spatial weights >

Exploratory spatial data analysis >

Point patterns >

Clustering >

Raster data >

Interpolation >

Regression >

Relationship and prediction >

Regression and geography

Fit that model

```

significant = np.abs(results.tvalues) > critical_t

fig, axs = plt.subplots(4, 3, figsize=(9, 9))
axs = axs.flatten()
for i, name in enumerate(independent_names[1:-1]):
    significant_mask = significant[:, i + 1]
    elections_data.plot(results.params[:, i + 1], cmap="plasma", ax=axs[i])
    elections_data[~significant_mask].plot(color="white", ax=axs[i], alpha=.9)
    axs[i].set_title(name[:20], fontdict={'fontsize': 8})
    axs[i].set_axis_off()

```

undetermined

incomplete_primary_e

lower_secondary_and_

further_education

post_maturita_studie

bachelors_degree

doctoral_degree

masters_degree

higher_vocational_ed

higher_vocational_ed

primary_education

complete_secondary_v

Local coefficients

It seems that the coefficients are significant only in some areas, so you need to be careful when drawing conclusions here. This can be due to a lack of relationship or a small sample size. Try for yourself how the significance changes if you increase the bandwidth. But be careful as too large a bandwidth may miss regional differences and a bandwidth that would cover the whole country would be equal to the OLS model.

Adaptive bandwidth

If you'd like to use the adaptive bandwidth, you can use same tools. Just specify `fixed=False` in both `Set BW`

On this page

Data

Non-spatial linear regression

Spatial heterogeneity

Geographically weighted regression

Fixed bandwidth

Report an issue

Other Formats

Jupyter





Martin Fleischmann, MSc, PhD

martinfleischmann.net

martin.fleischmann@natur.cuni.cz

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_



Úřad vlády
České republiky



TÝDEN PRO
Digitální
Česko

budoucnostjedigitalni.gov.cz