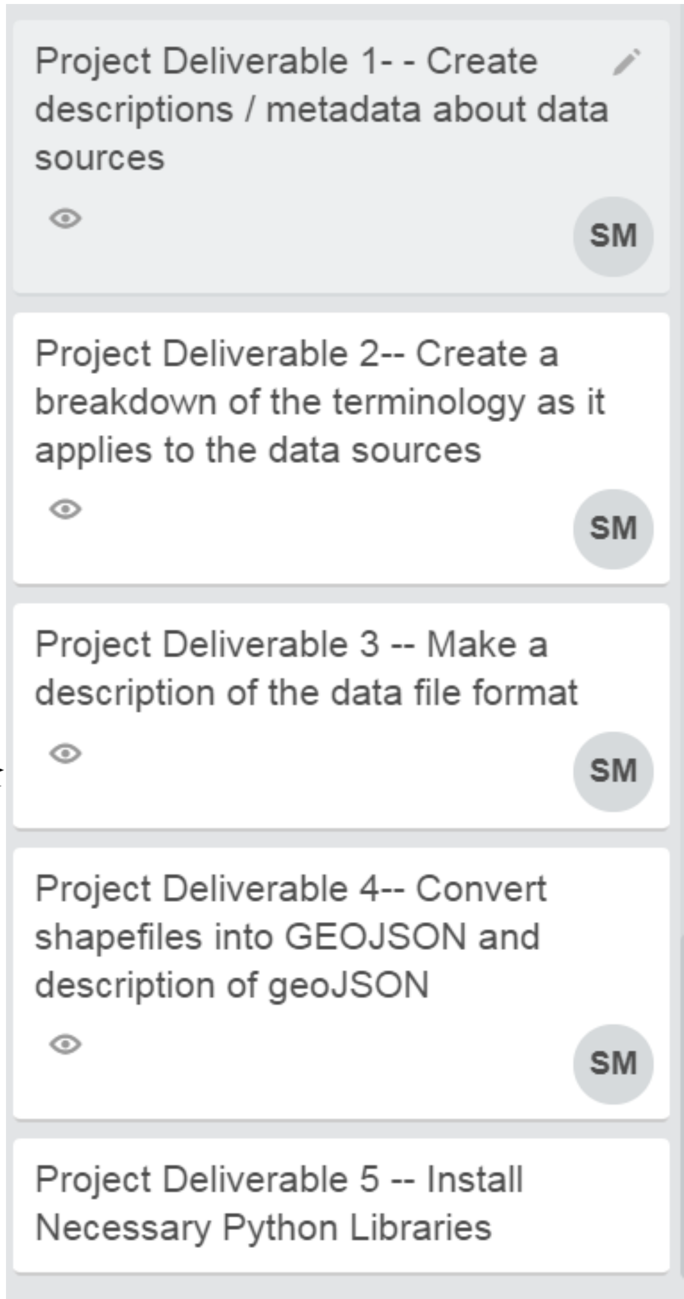
Sprint 1 Review

[Data Formats and Terminology]

OBJECTIVES:



Project Deliverable 1- - Create descriptions / metadata about data sources

Project Deliverable 2-- Create a breakdown of the terminology as it applies to the data sources

Project Deliverable 3 -- Make a description of the data file format

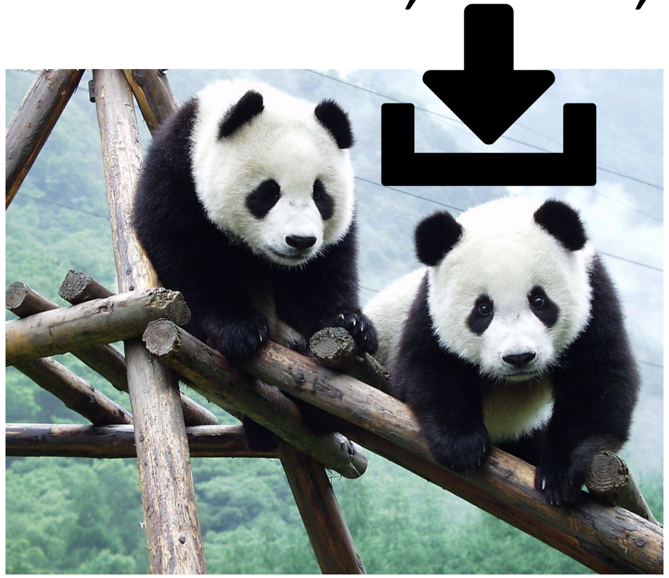
Project Deliverable 4-- Convert shapefiles into GEOJSON and description of geoJSON

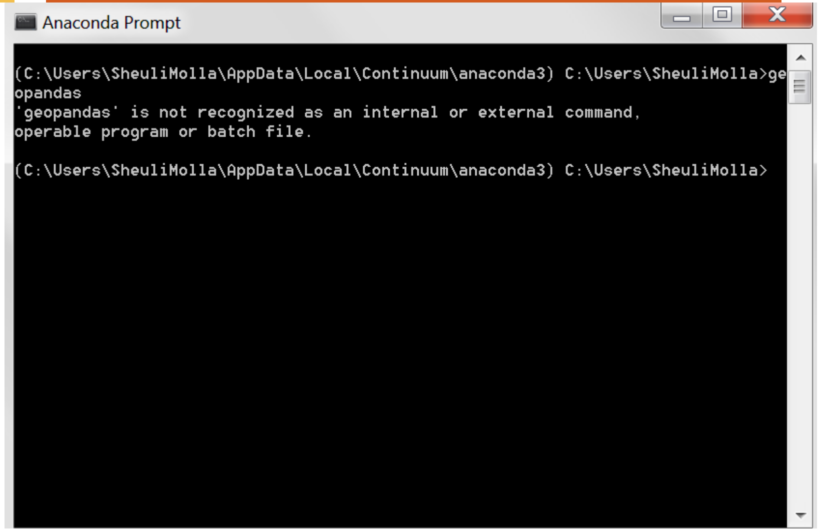
Project Deliverable 5 -- Install Necessary Python Libraries

QUESTIONS:

Q1. What kind of data sets do I need to reference to write environmental assessments? Q2. What terms do I need to know to work with geospatial files? Q3. What kind of data file formats do I see for the data sets I use regularly? Q4. How do I convert shape files to json or geojson? Q5. What libraries are most useful for geospatial work?

DOWNLOADING GEOSPATIAL LIBRARIES WITH ANACONDA PROMPT: GEOPANDAS, FIONA, ET AL.





```
(C:\Users\SheuliMolla\AppData\Local\Continuum\anaconda3) C:\Users\SheuliMolla>ge
opandas
'geopandas' is not recognized as an internal or external command,
operable program or batch file.

(C:\Users\SheuliMolla\AppData\Local\Continuum\anaconda3) C:\Users\SheuliMolla>
```

Installation via *conda* should also install all dependencies, but a complete list is as follows:

- numpy
- pandas (version 0.15.2 or later)
- shapely
- fiona
- six
- pyproj

```
conda install -c conda-forge geopandas
```

devleague

DATA FORMATS: Geospatial data is formatted in many ways.

RASTER

ADRG – National Geospatial-Intelligence Agency DRG) – digital scan of a paper USGS topographic map
ECRG – National Geospatial-Intelligence Agency (NGA)'s Enhanced Compressed ARC Raster Graphics (Better resolution than CADRG and no color loss)
ECW – Enhanced Compressed Wavelet (from ERDAS). A compressed wavelet format, often lossy.
Esri grid – proprietary binary and metadataless ASCII raster formats used by Esri
GeoTIFF – TIFF variant enriched with GIS relevant metadata
IMG – ERDAS IMAGINE image file format
JPEG2000 – Open-source raster format. A compressed format, allows both lossy and lossless compression.
MrSID – Multi-Resolution Seamless Image Database (by Lizardtech). A compressed wavelet format, allows both lossy and lossless compression.
netCDF-CF – netCDF file format with CF medata conventions for earth science data. Binary storage in open format with optional compression. Allows for direct web-access of subsets/aggregations of maps through OPeNDAPprotocol.
RPF – Raster Product Format, military file format specified in MIL-STD-2411[4]
CADRG – Compressed ADRG, developed by NGA, nominal compression of 55:1 over ADRG (type of Raster Product Format)**CIB** – Controlled Image Base, developed by NGA (type of Raster Product Format)

VECTOR

AutoCAD DXF – contour elevation plots in AutoCAD DXF format (by Autodesk)
Cartesian coordinate system (XYZ) – simple point cloud
Digital line graph (DLG) – a USGS format for vector data
Esri TIN - proprietary binary format for triangulated irregular network data used by Esri
Geography Markup Language (GML) – XML based open standard (by OpenGIS) for GIS data exchange
GeoJSON – a lightweight format based on JSON, used by many open source GIS packages
GeoMedia – Intergraph's Microsoft Access based format for spatial vector storage
ISFC – Intergraph's MicroStation based CAD solution attaching vector elements to a relational Microsoft Access database
Keyhole Markup Language (KML) – XML based open standard (by OpenGIS) for GIS data exchange
MapInfo TAB format – MapInfo's vector data format using TAB, DAT, ID and MAP files
National Transfer Format (NTF) – National Transfer Format (mostly used by the UK Ordnance Survey)
Spatialite – is a spatial extension to SQLite, providing vector geodatabase functionality. It is similar to PostGIS, Oracle Spatial, and SQL Server with spatial extensions
Shapefile – a popular vector data GIS format, developed by Esri
Simple Features – Open Geospatial Consortium specification for vector data
SOSI – a spatial data format used for all public exchange of spatial data in Norway
Spatial Data File – Autodesk's high-performance geodatabase format, native to MapGuide
TIGER – Topologically Integrated Geographic Encoding and Referencing
Vector Product Format (VPF) – National Geospatial-Intelligence Agency (NGA)'s format of vectored data for large geographic databases

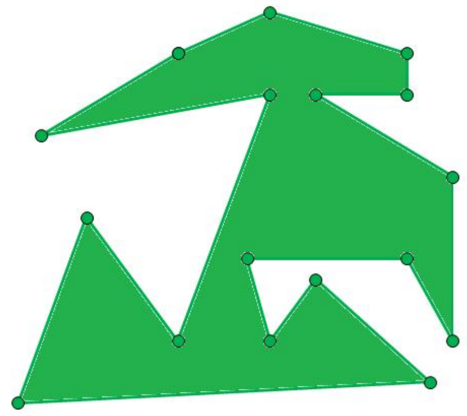
FORMATS BEST FORMATS FOR ME

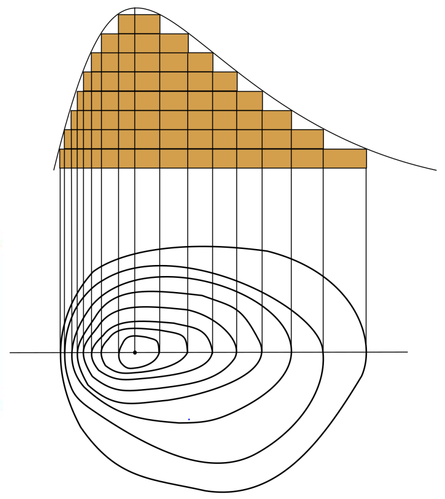
For my purposes, I am working almost exclusively with shapefiles (.shp), which were developed by ESRI for use in ArcGIS programs.

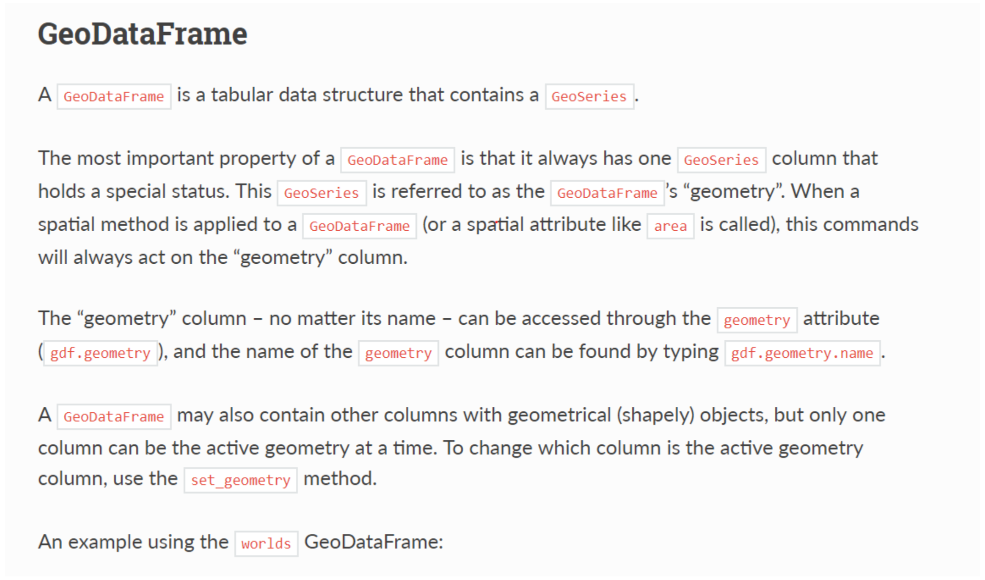
Similarly useful json and geojson files may be used with open source software (such as Mapbox).

BOTH Raster file types store the following information:

- LOCATION
- SHAPE
- ATTRIBUTES






GeoDataFrame

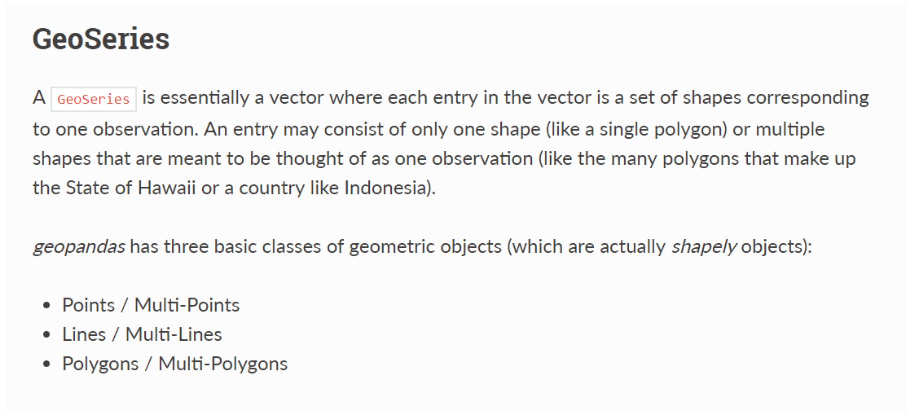
A `GeoDataFrame` is a tabular data structure that contains a `GeoSeries`.

The most important property of a `GeoDataFrame` is that it always has one `GeoSeries` column that holds a special status. This `GeoSeries` is referred to as the `GeoDataFrame`'s "geometry". When a spatial method is applied to a `GeoDataFrame` (or a spatial attribute like `area` is called), this commands will always act on the "geometry" column.

The "geometry" column – no matter its name – can be accessed through the `geometry` attribute (`gdf.geometry`), and the name of the `geometry` column can be found by typing `gdf.geometry.name`.

A `GeoDataFrame` may also contain other columns with geometrical (shapely) objects, but only one column can be the active geometry at a time. To change which column is the active geometry column, use the `set_geometry` method.

An example using the `worlds` GeoDataFrame:

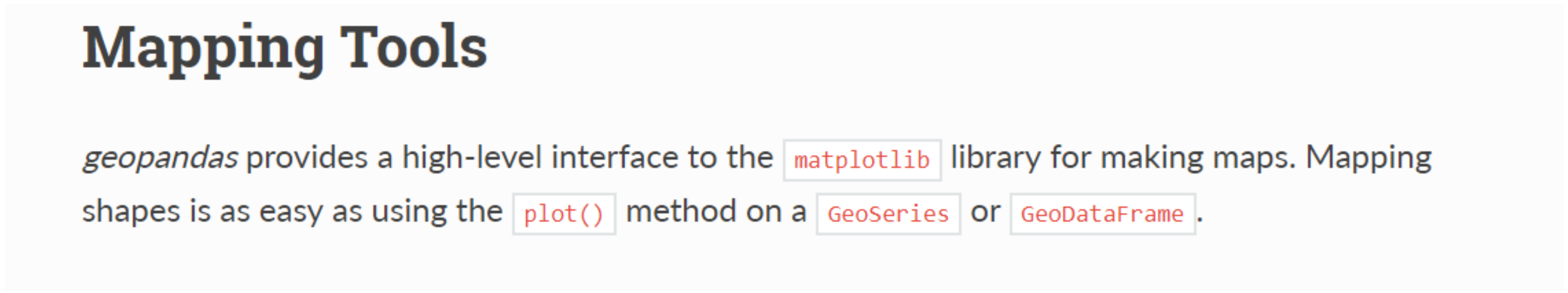

GeoSeries

A `GeoSeries` is essentially a vector where each entry in the vector is a set of shapes corresponding to one observation. An entry may consist of only one shape (like a single polygon) or multiple shapes that are meant to be thought of as one observation (like the many polygons that make up the State of Hawaii or a country like Indonesia).

geopandas has three basic classes of geometric objects (which are actually *shapely* objects):

- Points / Multi-Points
- Lines / Multi-Lines
- Polygons / Multi-Polygons

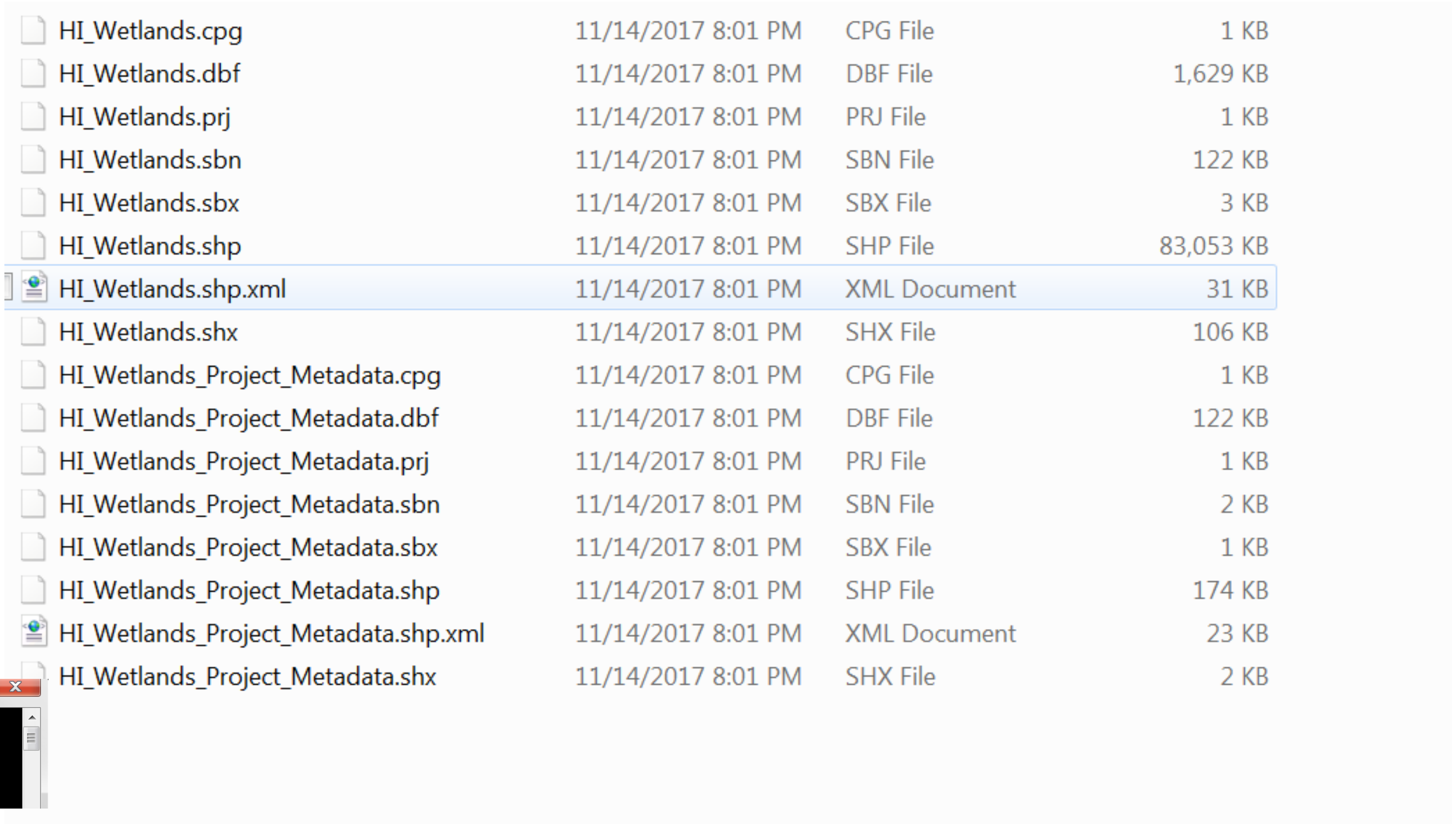
AFTER SOME TIME READING FROM THE GEOPANDAS WEBSITE…


Mapping Tools

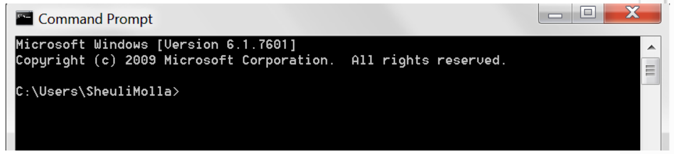
geopandas provides a high-level interface to the `matplotlib` library for making maps. Mapping shapes is as easy as using the `plot()` method on a `GeoSeries` or `GeoDataFrame`.

devleague

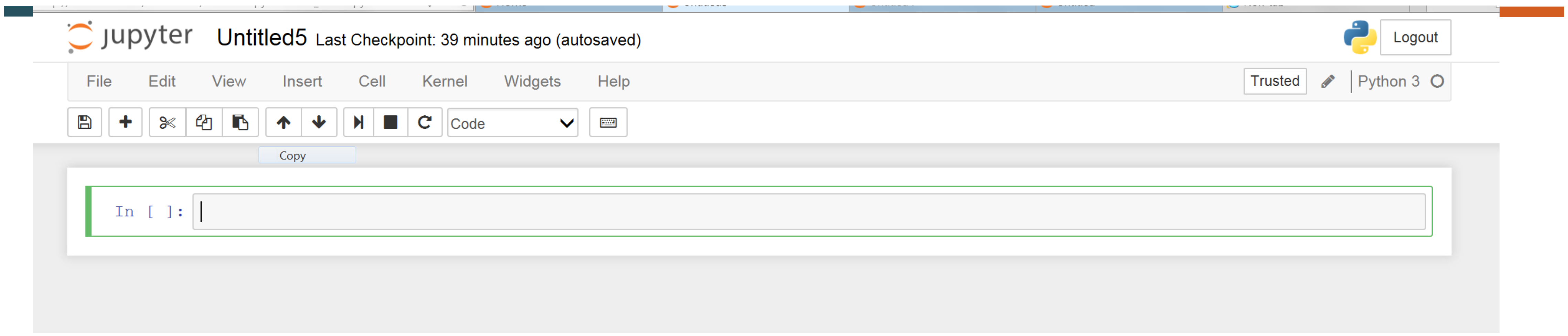
AND FUSSING WITH SOME WETLANDS DATA…



HI_Wetlands.cpg	11/14/2017 8:01 PM	CPG File	1 KB
HI_Wetlands.dbf	11/14/2017 8:01 PM	DBF File	1,629 KB
HI_Wetlands.prj	11/14/2017 8:01 PM	PRJ File	1 KB
HI_Wetlands.sbn	11/14/2017 8:01 PM	SBN File	122 KB
HI_Wetlands.sbx	11/14/2017 8:01 PM	SBX File	3 KB
HI_Wetlands.shp	11/14/2017 8:01 PM	SHP File	83,053 KB
HI_Wetlands.shp.xml	11/14/2017 8:01 PM	XML Document	31 KB
HI_Wetlands.shx	11/14/2017 8:01 PM	SHX File	106 KB
HI_Wetlands_Project_Metadata.cpg	11/14/2017 8:01 PM	CPG File	1 KB
HI_Wetlands_Project_Metadata.dbf	11/14/2017 8:01 PM	DBF File	122 KB
HI_Wetlands_Project_Metadata.prj	11/14/2017 8:01 PM	PRJ File	1 KB
HI_Wetlands_Project_Metadata.sbn	11/14/2017 8:01 PM	SBN File	2 KB
HI_Wetlands_Project_Metadata.sbx	11/14/2017 8:01 PM	SBX File	1 KB
HI_Wetlands_Project_Metadata.shp	11/14/2017 8:01 PM	SHP File	174 KB
HI_Wetlands_Project_Metadata.shp.xml	11/14/2017 8:01 PM	XML Document	23 KB
HI_Wetlands_Project_Metadata.shx	11/14/2017 8:01 PM	SHX File	2 KB

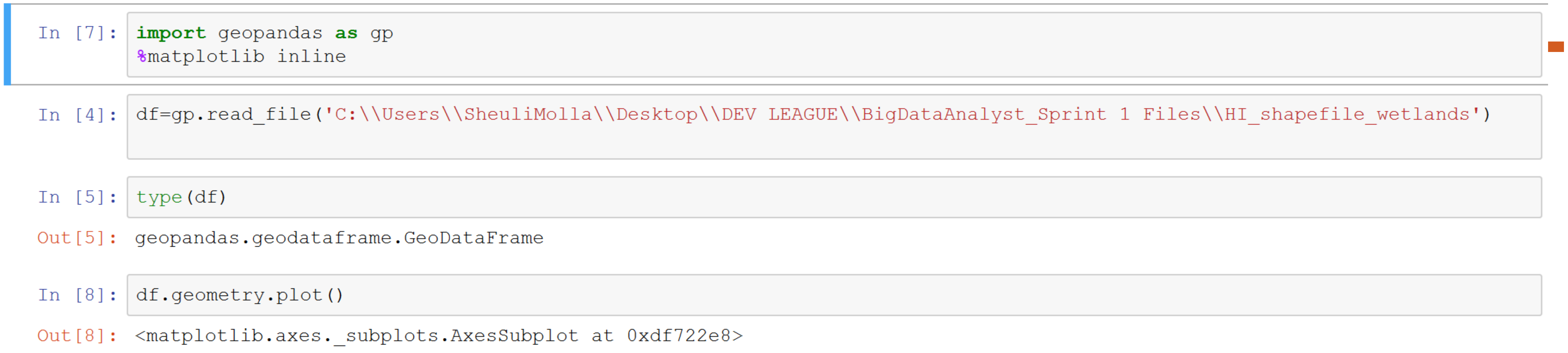


```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\SheuliMolla>
```

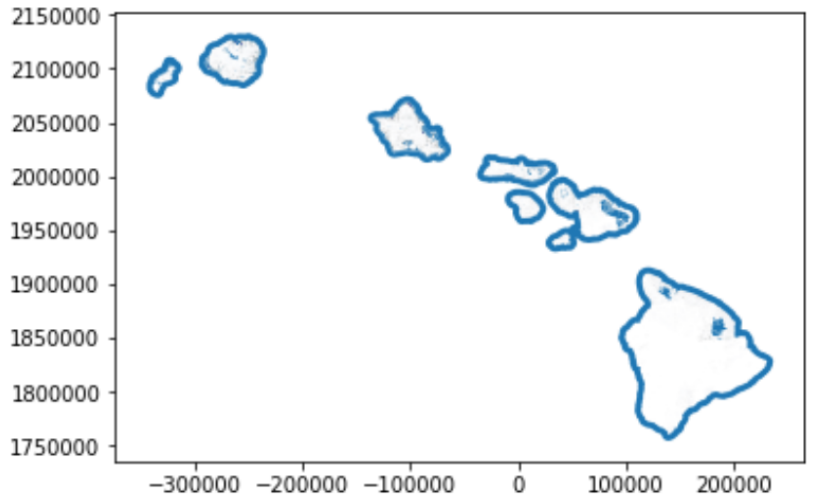


AND FIDDLING WITH MY JUPITER NOTEBOOK…

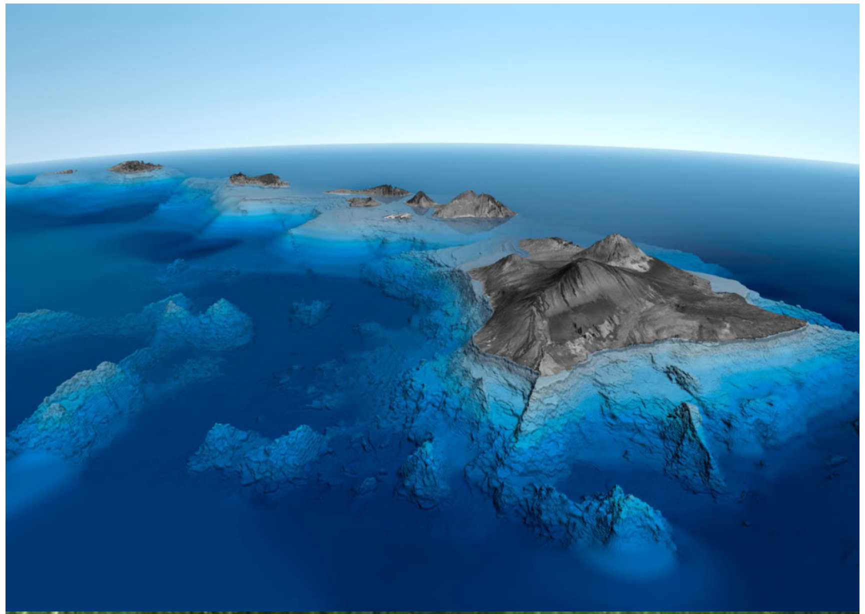
devleague



```
In [7]: import geopandas as gp
        %matplotlib inline

In [4]: df=gp.read_file('C:\\Users\\SheuliMolla\\Desktop\\DEV LEAGUE\\BigDataAnalyst_Sprint 1 Files\\HI_shapefile_wetlands')

In [5]: type(df)

Out[5]: geopandas.geodataframe.GeoDataFrame

In [8]: df.geometry.plot()

Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0xdf722e8>
```



…I MADE THIS



I IMPORTED A DATA FILE OF ALL OF HAWAII'S WETLANDS AND MAPPED THEM WITH LESS THAN 8 LINES OF CODE, AND I LEARNED A LOT IN THE PROCESS

Fin

devleague