module-57

October 31, 2024

1 Module 57: Storytelling with Data

Author: Juliho Castillo Colmenares Ph.D.

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file with the appropriate encoding due to decoding errors
file_path = "SYB66_153_202310_Gross Value Added by Economic Activity.csv"

data = pd.read_csv(file_path, encoding="ISO-8859-1")
data
```

[1]:
```
                      T14 Gross value added by kind of economic activity  \
0     Region/Country/Area                                             NaN
1                        4                                     Afghanistan
2                        4                                     Afghanistan
3                        4                                     Afghanistan
4                        4                                     Afghanistan
…                      …                                               …
4426                   716                                      Zimbabwe
4427                   716                                      Zimbabwe
4428                   716                                      Zimbabwe
4429                   716                                      Zimbabwe
4430                   716                                      Zimbabwe

     Unnamed: 2                                          Unnamed: 3 Unnamed: 4  \
0          Year                                              Series      Value
1          1995  Agriculture, hunting, forestry and fishing (% …       66.3
2          2005  Agriculture, hunting, forestry and fishing (% …       37.3
3          2010  Agriculture, hunting, forestry and fishing (% …       33.2
4          2015  Agriculture, hunting, forestry and fishing (% …       27.3
…         …                                                   …         …
4426       2010                 Services (% of gross value added)       65.6
4427       2015                 Services (% of gross value added)       65.7
4428       2019                 Services (% of gross value added)       55.7
4429       2020                 Services (% of gross value added)       56.0
4430       2021                 Services (% of gross value added)       60.1
```

```
                                      Unnamed: 5  \
0                                      Footnotes
1     Data classified according to ISIC Rev. 4.
2     Data classified according to ISIC Rev. 4.
3     Data classified according to ISIC Rev. 4.
4     Data classified according to ISIC Rev. 4.
…                                             …
4426                                        NaN
4427                                        NaN
4428                                        NaN
4429                                        NaN
4430                                        NaN

                                               Unnamed: 6
0                                                  Source
1     United Nations Statistics Division, New York, …
2     United Nations Statistics Division, New York, …
3     United Nations Statistics Division, New York, …
4     United Nations Statistics Division, New York, …
…                                                   …
4426  United Nations Statistics Division, New York, …
4427  United Nations Statistics Division, New York, …
4428  United Nations Statistics Division, New York, …
4429  United Nations Statistics Division, New York, …
4430  United Nations Statistics Division, New York, …

[4431 rows x 7 columns]
```

[2]:
```python
# Rename columns to improve readability
data.columns = [
    "Region/Country/Area",
    "Economic Activity",
    "Year",
    "Series",
    "Value",
    "Footnotes",
    "Source",
]

# Remove any rows where the 'Region/Country/Area' or 'Year' is empty or has␣
 ↪header-like content
data_cleaned = data.dropna(subset=["Region/Country/Area", "Year"]).reset_index(
    drop=True
)
```

```
# Filter out rows where 'Region/Country/Area' is actually "Region/Country/Area"␣
 ↪to remove initial header rows
data_cleaned = data_cleaned[
    data_cleaned["Region/Country/Area"] != "Region/Country/Area"
]

# Convert 'Year' and 'Value' columns to numeric, handling errors in conversion␣
 ↪to avoid issues with data type
data_cleaned["Year"] = pd.to_numeric(data_cleaned["Year"], errors="coerce")
data_cleaned["Value"] = pd.to_numeric(data_cleaned["Value"], errors="coerce")

# Drop any remaining rows with NaN values in 'Year' or 'Value' after conversion
data_cleaned = data_cleaned.dropna(subset=["Year", "Value"]).
 ↪reset_index(drop=True)
data_cleaned
```

[2]:

	Region/Country/Area	Economic Activity	Year	\
0	4	Afghanistan	1995	
1	4	Afghanistan	2005	
2	4	Afghanistan	2010	
3	4	Afghanistan	2015	
4	4	Afghanistan	2019	
…	…	…	…	
4425	716	Zimbabwe	2010	
4426	716	Zimbabwe	2015	
4427	716	Zimbabwe	2019	
4428	716	Zimbabwe	2020	
4429	716	Zimbabwe	2021	

	Series	Value	\
0	Agriculture, hunting, forestry and fishing (% …	66.3	
1	Agriculture, hunting, forestry and fishing (% …	37.3	
2	Agriculture, hunting, forestry and fishing (% …	33.2	
3	Agriculture, hunting, forestry and fishing (% …	27.3	
4	Agriculture, hunting, forestry and fishing (% …	27.0	
…	…	…	
4425	Services (% of gross value added)	65.6	
4426	Services (% of gross value added)	65.7	
4427	Services (% of gross value added)	55.7	
4428	Services (% of gross value added)	56.0	
4429	Services (% of gross value added)	60.1	

	Footnotes	\
0	Data classified according to ISIC Rev. 4.	
1	Data classified according to ISIC Rev. 4.	
2	Data classified according to ISIC Rev. 4.	
3	Data classified according to ISIC Rev. 4.	

```
4     Data classified according to ISIC Rev. 4.;Excl…
…                                                   …
4425                                                  NaN
4426                                                  NaN
4427                                                  NaN
4428                                                  NaN
4429                                                  NaN

                                                   Source
0     United Nations Statistics Division, New York, …
1     United Nations Statistics Division, New York, …
2     United Nations Statistics Division, New York, …
3     United Nations Statistics Division, New York, …
4     United Nations Statistics Division, New York, …
…                                                   …
4425  United Nations Statistics Division, New York, …
4426  United Nations Statistics Division, New York, …
4427  United Nations Statistics Division, New York, …
4428  United Nations Statistics Division, New York, …
4429  United Nations Statistics Division, New York, …

[4430 rows x 7 columns]
```

[3]:
```
# Filter for Afghanistan and a specific sector (Agriculture, hunting, forestry,␣
 ↪and fishing)
afghanistan_data = data_cleaned[
    (data_cleaned["Region/Country/Area"] == "4")
    & (
        data_cleaned["Series"].str.contains(
            "Agriculture, hunting, forestry and fishing"
        )
    )
]
afghanistan_data
```

[3]:
```
  Region/Country/Area Economic Activity  Year  \
0                   4       Afghanistan  1995
1                   4       Afghanistan  2005
2                   4       Afghanistan  2010
3                   4       Afghanistan  2015
4                   4       Afghanistan  2019
5                   4       Afghanistan  2020
6                   4       Afghanistan  2021

                                              Series  Value  \
0  Agriculture, hunting, forestry and fishing (% …   66.3
1  Agriculture, hunting, forestry and fishing (% …   37.3
```

```
2  Agriculture, hunting, forestry and fishing (% …   33.2
3  Agriculture, hunting, forestry and fishing (% …   27.3
4  Agriculture, hunting, forestry and fishing (% …   27.0
5  Agriculture, hunting, forestry and fishing (% …   28.1
6  Agriculture, hunting, forestry and fishing (% …   35.0

                                          Footnotes  \
0          Data classified according to ISIC Rev. 4.
1          Data classified according to ISIC Rev. 4.
2          Data classified according to ISIC Rev. 4.
3          Data classified according to ISIC Rev. 4.
4  Data classified according to ISIC Rev. 4.;Excl…
5  Data classified according to ISIC Rev. 4.;Excl…
6  Data classified according to ISIC Rev. 4.;Excl…

                                               Source
0  United Nations Statistics Division, New York, …
1  United Nations Statistics Division, New York, …
2  United Nations Statistics Division, New York, …
3  United Nations Statistics Division, New York, …
4  United Nations Statistics Division, New York, …
5  United Nations Statistics Division, New York, …
6  United Nations Statistics Division, New York, …
```

[4]:
```python
# Set up the figure and axis for the plot
fig, ax = plt.subplots(figsize=(10, 6))

# Plotting the data with Tufte's principle of simplicity by reducing grid lines
ax.plot(
    afghanistan_data["Year"],
    afghanistan_data["Value"],
    marker="o",
    linestyle="-",
    color="royalblue",
    label="Agriculture, Forestry, and Fishing",
)

# Gestalt: Grouping elements to indicate continuity in time by using a line plot
# Tufte: Using minimalist axis and labels to focus on data
ax.set_title("Agricultural Value Added in Afghanistan (1995-2019)", fontsize=14)
ax.set_xlabel("Year", fontsize=12)
ax.set_ylabel("Value Added (%)", fontsize=12)
ax.grid(True, which="both", linestyle="--", linewidth=0.5, color="gray",␣
 ↪alpha=0.7)

# Holmes: Using color and annotations for contrast and clarity without␣
 ↪overwhelming color usage
```

```
ax.legend()
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)

# Final annotations and displaying the plot
plt.tight_layout()
plt.show()
```