

# Progetto di Data Viz


- Questo progetto serve a verificare il vostro **grado di comprensione** del modulo di **data visualization**
- Ulteriore obiettivo di questo compito è quello di rafforzare le **vostre competenze tecniche**: vi viene chiesto in particolare di realizzare un vero e proprio **sistema di dataviz** a partire da un **dataset** reale
- Il progetto può essere sviluppato in gruppi da 1-2 persone.

# Il dataset

- Per prima cosa dovete scegliere una sorgente di dati da questo portale:
  - <http://data.un.org/Default.aspx>
- Si tratta del progetto di apertura dei dati delle Nazioni Unite
- Selezionare un database a scelta

Explorer

Datasets Sources Topics

- + Commodity Trade Statistics Database United Nations Statistics Division (UNSD) ⓘ
- + Energy Statistics Database United Nations Statistics Division (UNSD) ⓘ
- + Environment Statistics Database United Nations Statistics Division (UNSD) ⓘ
- + FAO Data Food and Agriculture Organization (FAO) ⓘ
- + Gender Info United Nations Statistics Division (UNSD) ⓘ
- + Global Indicator Database United Nations Statistics Division (UNSD) ⓘ
- + Greenhouse Gas Inventory Data United Nations Framework Convention on Climate Change (UNFCCC) ⓘ
- + Human Development Indices: A statistical update 2015 United Nations Development Programme (UNDP) ⓘ
- + Indicators on Women and Men United Nations Statistics Division (UNSD) ⓘ
- + INDSTAT United Nations Industrial Development Organization (UNIDO) ⓘ
- + Industrial Commodity Statistics Database United Nations Statistics Division (UNSD) ⓘ
- + International Financial Statistics International Monetary Fund (IMF) ⓘ
- Key Indicators of the Labour Market, 7th Edition International Labour Organization (ILO) ⓘ
  - Youth unemployment, both sexes  [View data](#)
  - Youth unemployment, men [View data](#)
  - Youth unemployment, women [View data](#)
- + LABORSTA International Labour Organization (ILO) ⓘ
- + Millennium Development Goals Database United Nations Statistics Division (UNSD) ⓘ ⚡
- + National Accounts Estimates of Main Aggregates United Nations Statistics Division (UNSD) ⓘ
- + National Accounts Official Country Data United Nations Statistics Division (UNSD) ⓘ
- + OECD Data Organisation for Economic Co-operation and Development (OECD) ⓘ
- + The State of the World's Children United Nations Children's Fund (UNICEF) ⓘ
- + UIS Data Centre UNESCO Institute for Statistics (UNESCO UIS) ⓘ ⚡

Youth unemployment, both sexes [Search glossaries](#)Source: Key Indicators of the Labour Market, 7th Edition | [International Labour Organization](#)[Download](#) [Explore](#)

Country	Year	Sex	Youth unemployed ('000)	Youth labour force ('000)	Youth unemployment rate (%)	Adult unemployed ('000)	Adult labour force ('000)	Ratio of youth unemployment rate to adult unemployment rate	Total unemployed ('000)	Share of youth unemployed in total unemployed (%)	Youth population ('000)	Share of youth unemployed in youth population (%)	Adult unemployment rate (%)
Albania	2001	MF	111.799	315.3	35.5	193.707	1 031.8	1.9	305.506	36.6	527.4	21.2	18.8
Albania	2007 b	MF			20.1			1.7					12.0
Albania	2008	MF			27.2			2.5					10.9
Albania	2009	MF			27.2			2.4					11.3
Algeria	2001	MF	1 081.4	2 264.3	47.8	1 258.1	6 304.3	2.4	2 339.5	46.2			20.0
Algeria	2003	MF	996.0			1 082.3			2 078.3	47.9			
Algeria	2004	MF	762.3	1 756.3	43.4	909.2	6 564.1	3.1	1 671.5	45.6			13.9
Algeria	2005	MF	699.1	2 249.6	31.1	775.5	7 406.4	3.0	1 474.6	47.4			10.5
Algeria	2006	MF	536.4	2 207.5	24.3	704.4	7 902.2	2.7	1 240.8	43.2			8.9
American Samoa	1990	MF	0.31	2.6	11.9	0.416	11.9	3.4	0.726	42.7	8.5	3.6	3.5
Anguilla	1992	MF	0.18	1.1	16.4	0.16	3.35	3.4	0.34	52.9			4.8
Anguilla	2001 b	MF	0.149	1.118	13.3	0.248	4.931	2.6	0.397	37.5	1.756	8.5	5.0
Antigua and Barbuda	1991	MF	0.845	6.445	13.1	0.94	23.2	3.2	1.78	47.5			4.1
Antigua and Barbuda	2001 b	MF	1.253	6.3	19.9	2.075	33.1	3.2	3.328	37.7	12.4	10.1	6.3
Argentina	1980	MF	41.1			41.1			82.2	50.0			

# Ponetevi una domanda

- Deve essere semplice, ma intrigante
- Ad esempio:
  - “è vero che i paesi più ricchi sono quelli con il maggiore tasso di occupazione?”
  - “È vero che più gli adulti lavorano fino a tarda età, più è difficile per i giovani trovare lavoro?”
  - “È vero che oggi è più difficile trovare lavoro di qualche anno fa?”

# Studiate il database

- Il db selezionato sarà pieno di dati, colonne, campi.
- Cercate di **capire il significato dei valori**
- Il sito fornisce alcune informazioni di base, altre dovrete cercarle altrove, per avere almeno un'idea degli indicatori scelti, del dominio, se i valori devono essere normalizzati, etc.

# Che tipi di dati avete scelto?

- Con che tipo di dati avete a che fare?
  - Multivariati
  - 1d, 2d o 3d
  - Geo-spaziali
  - Temporal
  - Gerarchici
  - Etc.

# Plotting preliminare

- Iniziate a visualizzare qualcosa, in base al tipo di dato con il quale avete a che fare
- **Osservate il risultato**, provate a vedere se trovate un **segnale** della risposta alla domanda che vi siete posti
- Non preoccupatevi se i primi risultati sono “brutti”: **ripete iterativamente il processo**
  - Raffinate eventualmente la domanda
  - Fate plot di situazioni diverse
  - Confrontate i plot (magari con small multiples, scatter plot matrix, etc)



# Progettate il vostro sistema

- **Quale sistema** può aiutare meglio a trovare risposta alla domanda?
- L'utente deve **esplorare i dati** da solo: formulerà le sue istanze di domanda, dovrà trovare le sue risposte
  - Es. visti i pattern generali di occupazione, quali i dettagli dell'Italia nel 2010? O quelli relativi al resto d'Europa? O del mondo?
  - Non potete conoscere l'istanza particolare della domanda dell'utente a priori
  - Come potete aiutare l'utente ad esplorare meglio i dati? (overview+detail, focus+context, ...)
- Quali capacità *“pre-attentive”* possono aiutare meglio a focalizzare **l'attenzione dell'utente** sui particolari utili alla comprensione del dato?
- Attenzione agli **inganni percettivi**
- Problemi da risolvere **ad-hoc**: overplotting, animazione, interazione, etc.

# Realizzate il prototipo del vostro sistema

- Abbiamo visto viz fantastiche, alcune delle quali richiedono mesi di lavoro e gruppi molto compositi
  - Non vi è consentito alcun complesso del brutto anatroccolo!
- Pensate ad un lavoro che possa essere svolto in **due/tre settimane** di attività
- Potete usare **Python/plot.ly**, **d3.js** – altre librerie se preferite (anche R, Processing, etc.)

# Consegna

- **Scadenza:** la presentazione è da portare il giorno stesso dell'appello scelto
- Preparate una presentazione tipo power point per poter spiegare quello che avete fatto in 10 minuti
  - Dataset scelto
  - Domanda principale
  - Metodologia usata per progettare e realizzare il sistema
  - Demo

# Consiglio finale

- Prendete ispirazione da Hans Rosling e dal suo team (e da tanti altri ...)

***“Let my dataset change your mindset”***

**GAPMINDER**

a fact-based worldview

<http://www.gapminder.org/>