

Day in Life of Data Viz Engineer

Women in Big Data

Kat Gordienko
July 18, 2018

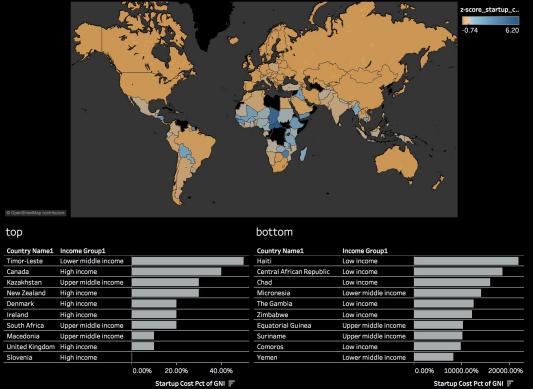
NETFLIX

Agenda

- **Presentation (~10 mins)**
- **Hands-on Tutorial (~50 mins)**

Data Visualization in One Slide

Country Name	Country Code	Indicator Name	Indicator Code	data
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.ZS	37.16521072	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.FE.ZS	25.63540268	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.MA.ZS	48.32851791	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.LO.ZS	42.54204559	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.AO.ZS	27.72478104	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.PL.ZS	26.45811081	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.G0.ZS	43.44695282	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTL.S0.ZS	48.66697693	
Arab World	ARB	Account ownership at a financial instit FX.OWN.TOTLYG.ZS	20.95479965	
Arab World	ARB	Arms exports (SIPRI trend indicator val MS.MILXPR.TD)	15600000	
Arab World	ARB	Arms imports (SIPRI trend indicator va MS.MILMPRT.KD)	11535000000	
Arab World	ARB	Bird species, threatened EN.BIR.THRD.NO	310	
Arab World	ARB	Burden of customs procedure, WEF (1:IQ.WEC.CUST.XQ)	4.085714286	
Arab World	ARB	Business extent of disclosure index (0+IC.BUS.DISC.XQ)	5.863636364	
Arab World	ARB	Contributing family workers, female (%SLFAM.WORK.FE.ZS)	18.94171494	
Arab World	ARB	Contributing family workers, male (%SLFAM.WORK.MA.ZS)	4.549224068	
Arab World	ARB	Contributing family workers, total (%SLFAM.WORK.ZS)	7.241652691	
Arab World	ARB	Cost of business start-up procedures (!IC.REG.COST.PC.ZS)	31.86363636	
Arab World	ARB	Cost of business start-up procedures, !IC.REG.COST.PC.FE.ZS	31.86363636	
Arab World	ARB	Cost of business start-up procedures, !IC.REG.COST.PC.MA.ZS	31.86363636	
Arab World	ARB	Cost to export, border compliance (US) CEXP.CSCB.CD	489.9545455	
Arab World	ARB	Cost to export, documentary compliant (US) CEXP.CSCD.CD	248.6818182	
Arab World	ARB	Cost to import, border compliance (US) CIIMP.CSCB.CD	580.6262626	
Arab World	ARB	Cost to import, border compliance (US) CIIMP.CSCD.CD	276.0454545	
Arab World	ARB	Depth of credit information index (0=IC.CRD.INFO.XQ)	4.045454545	

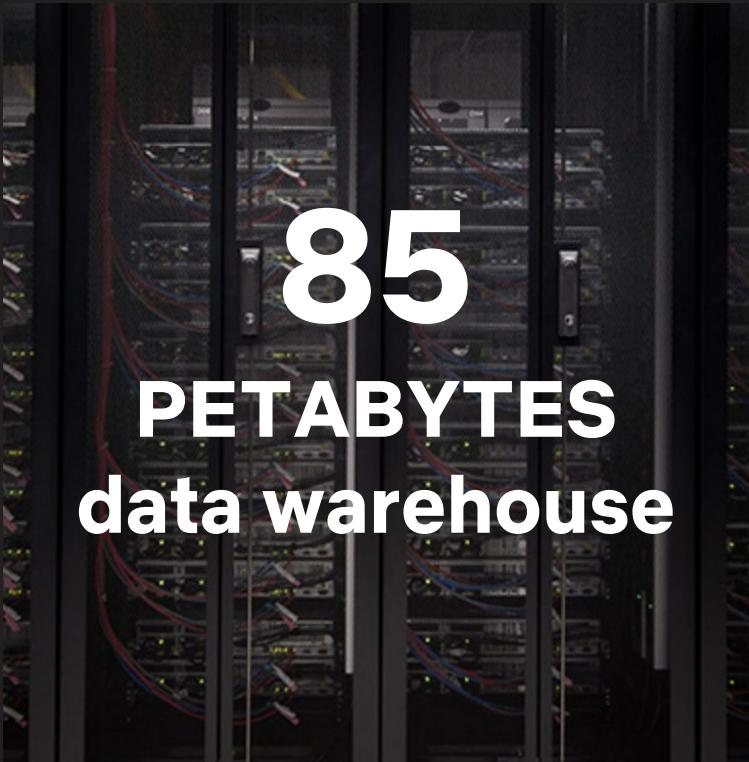


Raw data: processing, cleansing, mining, querying, aggregation

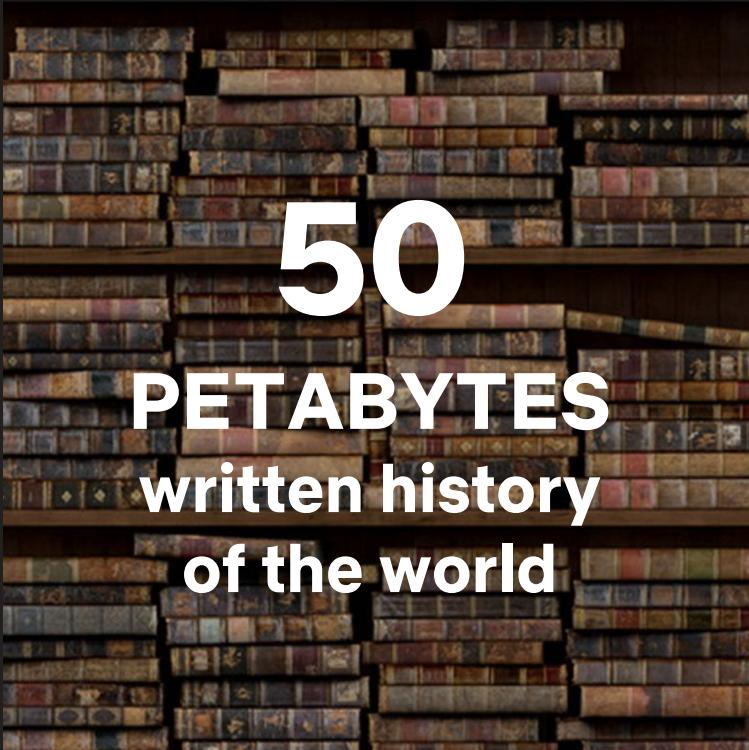
Analytics: business logic, KPI, insights, trends and patterns

Presentation: visual storytelling via reports and dashboards

Data @ Netflix

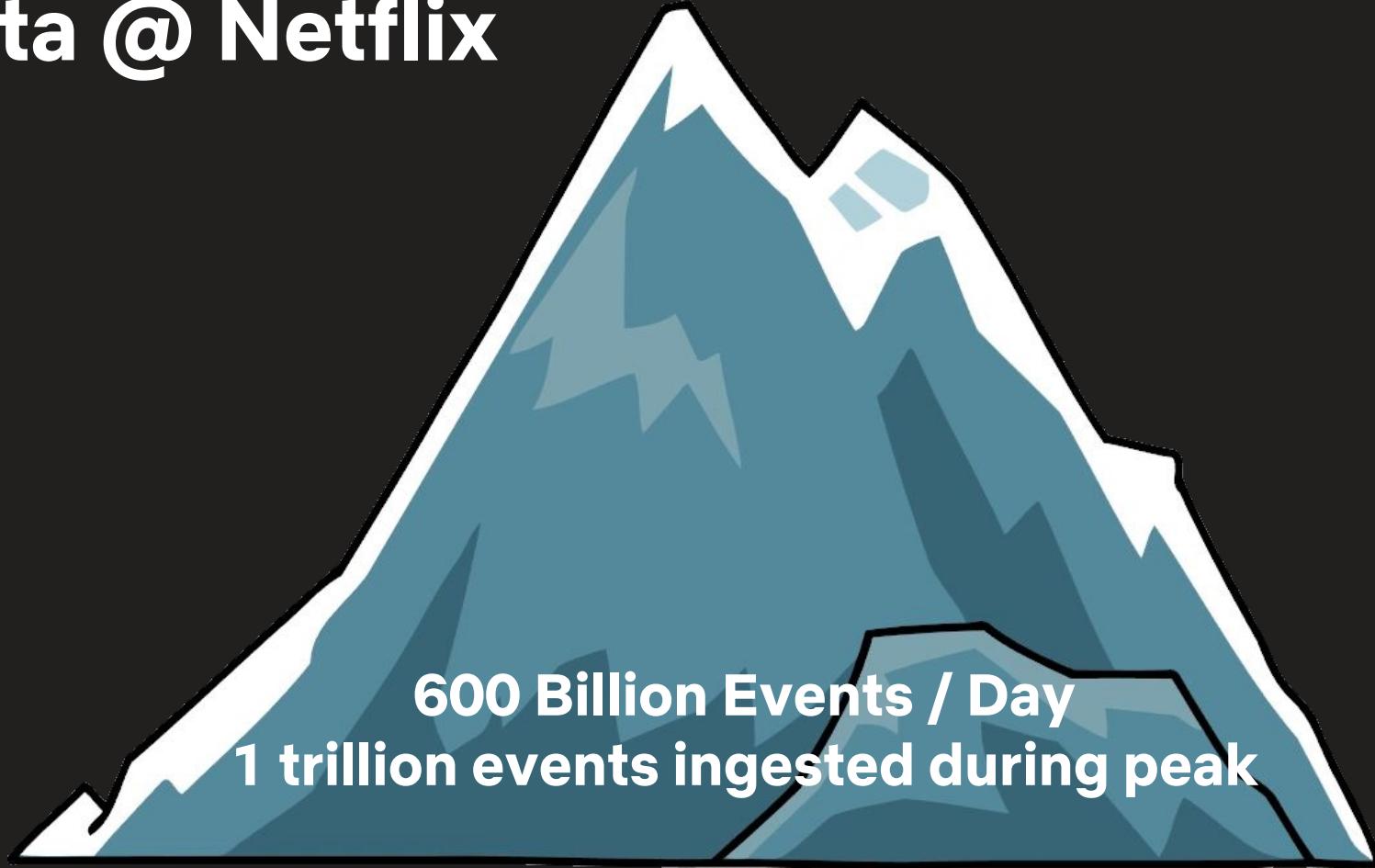


185
PETABYTES
data warehouse



50
PETABYTES
written history
of the world

Data @ Netflix



Analytics @ Netflix

Cross-functional collaboration between:

- Business Partners
- Product Managers
- Analytics Engineers
- Data Scientists
- Data Engineers
- Data Platform

Visuals @ Netflix

- >6,000 reports
- ~1,500 users



Typical Data Viz Project

1. Start with **business questions**:

- What is the business need and use cases?
- What decisions are going to be influenced?
- Who will be using the data product (analyst, exec)?
- What are the main KPIs to include in the product?
- What data sources should be included?
- What is the reporting granularity and timeframe?
- How much information should be included to provide enough detail?

Typical Data Viz Project

2. Evaluate **data**:

- Find data sources (internal, vendor, public)
- Understand schema and relationships
- Learn business rules and logic
- Prepare data sources for reporting

3. Come up with **visuals**:

- Pick the best tool for use case
- Analyze data, find patterns, define KPIs
- Create visuals to best represent the data

Typical Data Viz Project

4. Create **narrative**:

- Tell a compelling story
- Explain with visuals, narrate with words
- Provide context



Visualization Tools



MicroStrategy®



IMMUTABLE

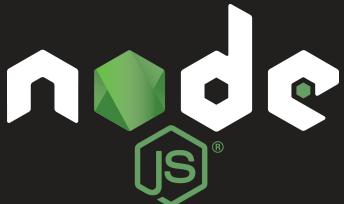
{less}

ES2015



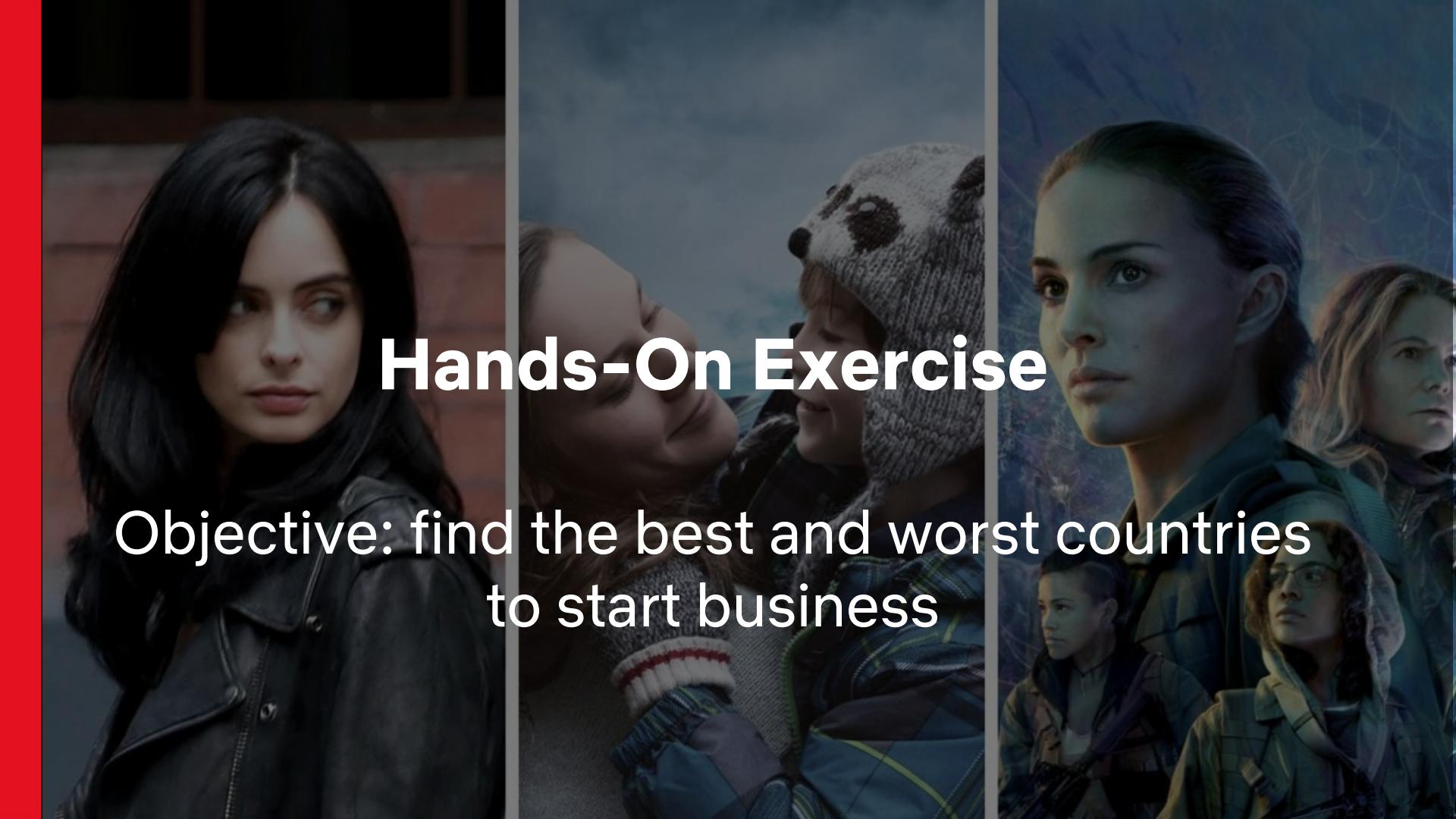
webpack

BABEL



express

npm



Hands-On Exercise

Objective: find the best and worst countries
to start business

What Is Tableau?

Tableau is business intelligence software that helps people see and understand their data.



Fast Analytics

Connect and visualize your data in minutes. Tableau is 10 to 100x faster than existing solutions.



Ease of Use

Anyone can analyze data with intuitive drag & drop products. No programming, just insight.



Big Data, Any Data

From spreadsheets to databases to Hadoop to cloud services, explore any data.



Smart Dashboards

Combine multiple views of data to get richer insight. Best practices of data visualization are baked right in.



Update Automatically

Get the freshest data with a live connection to your data or get automatic updates on a schedule you define.



Share in Seconds

Publish a dashboard with a few clicks to share it live on the web and on mobile devices.

Data Source

We'll be using the World Development Indicators Dataset published by World Bank.

Acronym:	WDI
Type:	Time Series
Topics:	Agriculture and Food Security, Climate Change, Economic Growth, Education, Energy and Extractives, Environment and Natural Resources, Financial Sector Development, Gender, Health, Nutrition and Population, Macroeconomic Vulnerability and Debt, Poverty, Private Sector Development, Public Sector Management, Social Development, Social Protection and Labor, Trade, Urban Development
Economy Coverage:	High Income, IBRD, IDA, Low Income, Lower Middle Income, Upper Middle Income
Languages Supported:	English, Arabic, Chinese, French, Spanish
Number of Economies:	217
Geographical Coverage:	World, East Asia & Pacific, American Samoa, Australia, Brunei Darussalam, Cambodia, China More...
External Contact Email:	Data@worldbank.org
Access Options:	Query Tool, API Documentation, Download
Temporal Coverage:	1960 - 2017
Update Frequency:	Quarterly



Instructions

- Go to <https://tinyurl.com/NFLXWIBD2018>
- Click on Data Viz folder
- Download viz_data.zip
- Extract downloaded zip file
- Open a new Tableau file

Z-Score (Standard Score)

The number of standard deviations from the mean a data point is.

$$z = (x - \mu) / \sigma$$

where:

μ is the mean of the population

σ is the standard deviation of the population

Summary

- Joins
- Top/bottom filters
- Maps
- Absolute values vs indices
- Index calculation (Z-Score)
- Calculations: Window and Level of Detail
- Parameters

THANK YOU!

kgordiienko@netflix.com

 /katgord

