World Development Indicators Project

Country Data - Data frame Definition

```scala
%scala

val country = sqlContext.read.format("csv")
  .option("header", "true")
  .option("inferSchema", "true")
  .load("/FileStore/tables/Country-3.csv")

display(country)
```

▶ (3) Spark Jobs

▶ country: org.apache.spark.sql.DataFrame = [CountryCode: string, ShortName: string ... 29 more fields]

Table ⌄ +

	CountryCode	ShortName	TableName	LongName	Alph
1	AFG	Afghanistan	Afghanistan	Islamic State of Afghanistan	AF
2	ALB	Albania	Albania	Republic of Albania	AL
3	DZA	Algeria	Algeria	People's Democratic Republic of Algeria	DZ
4	ASM	American Samoa	American Samoa	American Samoa	AS
5	ADO	Andorra	Andorra	Principality of Andorra	AD
6	AGO	Angola	Angola	People's Republic of Angola	AO
7	ATG	Antigua and Barbuda	Antigua and Barbuda	Antigua and Barbuda	AG
8	ARB	Arab World	Arab World	Arab World	1A
9	ARG	Argentina	Argentina	Argentine Republic	AR
10	ARM	Armenia	Armenia	Republic of Armenia	AM
11	ABW	Aruba	Aruba	Aruba	AW

⤓ 247 rows | 43.80 seconds runtime   Refreshed 33 days ago

Cmd 3

Create Temp View of Country Data

```scala
%scala

country.createOrReplaceTempView("country")
```

Cmd 4

```scala
%scala
country.printSchema()
```

```
root
 |-- CountryCode: string (nullable = true)
 |-- ShortName: string (nullable = true)
 |-- TableName: string (nullable = true)
 |-- LongName: string (nullable = true)
 |-- Alpha2Code: string (nullable = true)
 |-- CurrencyUnit: string (nullable = true)
 |-- SpecialNotes: string (nullable = true)
 |-- Region: string (nullable = true)
 |-- IncomeGroup: string (nullable = true)
 |-- Wb2Code: string (nullable = true)
 |-- NationalAccountsBaseYear: string (nullable = true)
 |-- NationalAccountsReferenceYear: string (nullable = true)
 |-- SnaPriceValuation: string (nullable = true)
 |-- LendingCategory: string (nullable = true)
 |-- OtherGroups: string (nullable = true)
 |-- SystemOfNationalAccounts: string (nullable = true)
 |-- AlternativeConversionFactor: string (nullable = true)
 |-- PppSurveyYear: string (nullable = true)
 |-- BalanceOfPaymentsManualInUse: string (nullable = true)
 |-- ExternalDebtReportingStatus: string (nullable = true)
```

```sql
%sql
select * from country;
```

▸ (1) Spark Jobs

Table ⌄ +

	CountryCode	ShortName	TableName	LongName	Alph
1	AFG	Afghanistan	Afghanistan	Islamic State of Afghanistan	AF
2	ALB	Albania	Albania	Republic of Albania	AL
3	DZA	Algeria	Algeria	People's Democratic Republic of Algeria	DZ
4	ASM	American Samoa	American Samoa	American Samoa	AS
5	ADO	Andorra	Andorra	Principality of Andorra	AD
6	AGO	Angola	Angola	People's Republic of Angola	AO
7	ATG	Antigua and Barbuda	Antigua and Barbuda	Antigua and Barbuda	AG
8	ARB	Arab World	Arab World	Arab World	1A
9	ARG	Argentina	Argentina	Argentine Republic	AR
10	ARM	Armenia	Armenia	Republic of Armenia	AM
11	ABW	Aruba	Aruba	Aruba	AW
12	AUS	Australia	Australia	Commonwealth of Australia	AU

Cmd 6

Indicators Data - Data frame Definition

Scala ▶⌄ ⊾⊿ ⌄ — ✕

```scala
%scala

val Indicators = sqlContext.read.format("csv")
  .option("header", "true")
  .option("inferSchema", "true")
  .load("/FileStore/tables/Indicators.csv")

display(Indicators)
```

▸ (3) Spark Jobs

▸ 🔲 Indicators: org.apache.spark.sql.DataFrame = [CountryName: string, CountryCode: string ... 4 more fields]

Table ⌄ +

	CountryName	CountryCode	IndicatorName	IndicatorCode
1	Arab World	ARB	Adolescent fertility rate (births per 1,000 women ages 15-19)	SP.ADO.TFRT
2	Arab World	ARB	Age dependency ratio (% of working-age population)	SP.POP.DPND
3	Arab World	ARB	Age dependency ratio, old (% of working-age population)	SP.POP.DPND.OI
4	Arab World	ARB	Age dependency ratio, young (% of working-age population)	SP.POP.DPND.YC
5	Arab World	ARB	Arms exports (SIPRI trend indicator values)	MS.MIL.XPRT.KD
6	Arab World	ARB	Arms imports (SIPRI trend indicator values)	MS.MIL.MPRT.KL
7	Arab World	ARB	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN
8	Arab World	ARB	CO2 emissions (kt)	EN.ATM.CO2E.KT
9	Arab World	ARB	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC
10	Arab World	ARB	CO2 emissions from gaseous fuel consumption (% of total)	EN.ATM.CO2E.GF
11				

⤓ ⌄ | 1,000 rows | Truncated data ⌄ | 41.51 seconds runtime   Refreshed 33 days ago

Cmd 7

Create Temp View on Indicators Data

```scala
%scala
Indicators.createOrReplaceTempView("Indicators")
```

Cmd 8

What is the Gini Index?

The Gini index is a simple measure of the distribution of income across income percentiles in a population. A higher Gini index indicates greater inequality, with high income individuals receiving much larger percentages of the total income of the population.
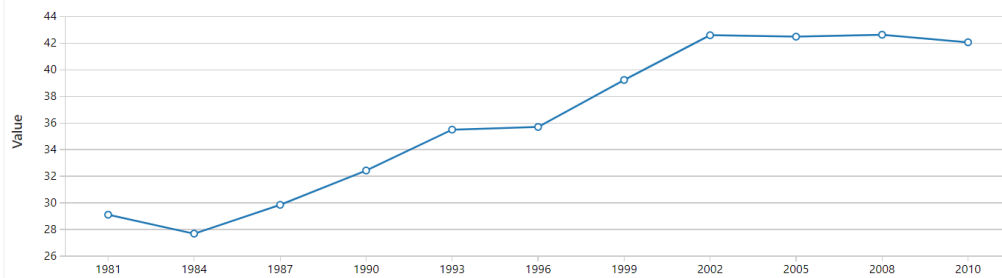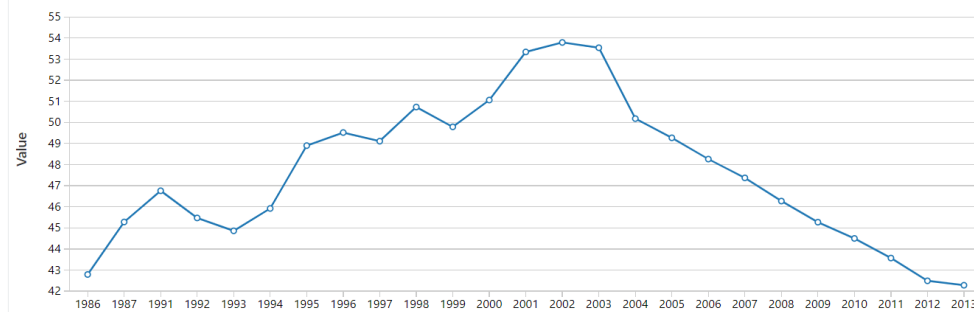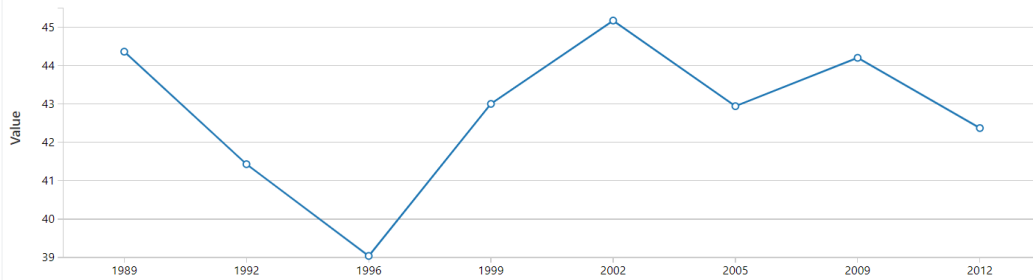
Cmd 9

GINI Index of China

SQL

```sql
%sql

select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "CHN" order by Year asc;
```

▼ (1) Spark Jobs
  ▼ Job 11 View (Stages: 1/1)
    Stage 11: 8/8 ⓘ

Visualization ∨   +

▶ (1) Spark Jobs

⬇ ✎ Edit Visualization 11 rows | 20.92 seconds runtime   Refreshed 33 days ago

GINI Index of Argentina

SQL

```sql
%sql

select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "ARG" order by Year asc;
```

▶ (1) Spark Jobs
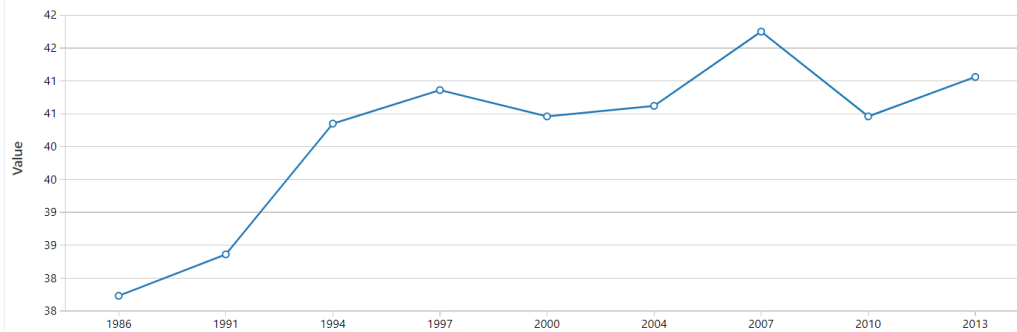
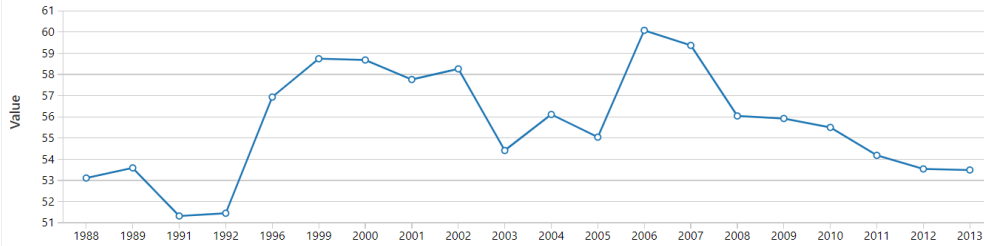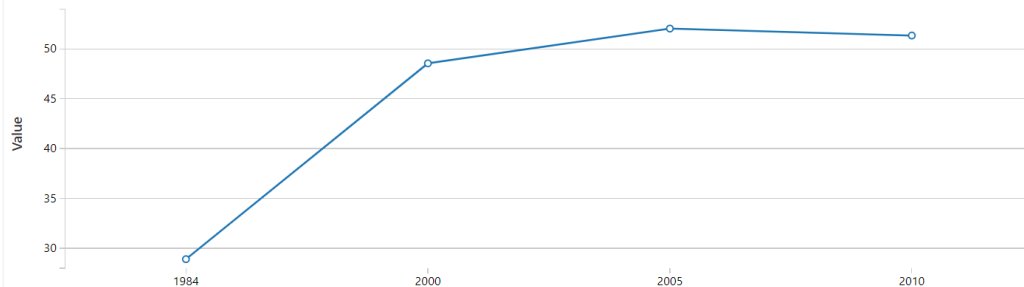Visualization ∨   +

▶ (1) Spark Jobs

GINI Index of Uganda

```
1   %sql
2
3   select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "UGA" order by Year asc;
```

▸ (1) Spark Jobs
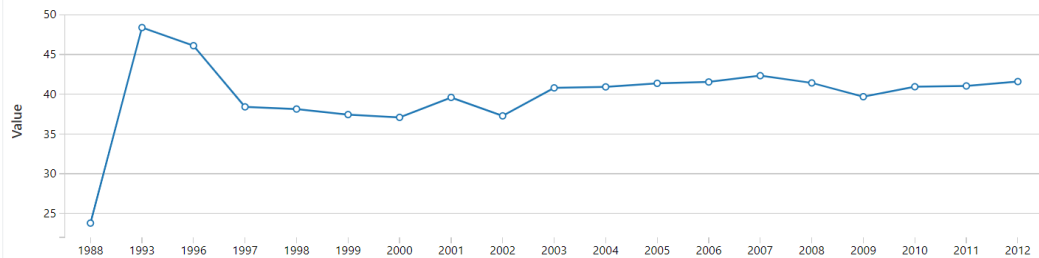
Visualization ⌄   +

▸ (1) Spark Jobs

GINI Index of USA

```
1   %sql
2
3   select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "USA" order by Year asc;
```

▸ (1) Spark Jobs

Visualization ⌄   +

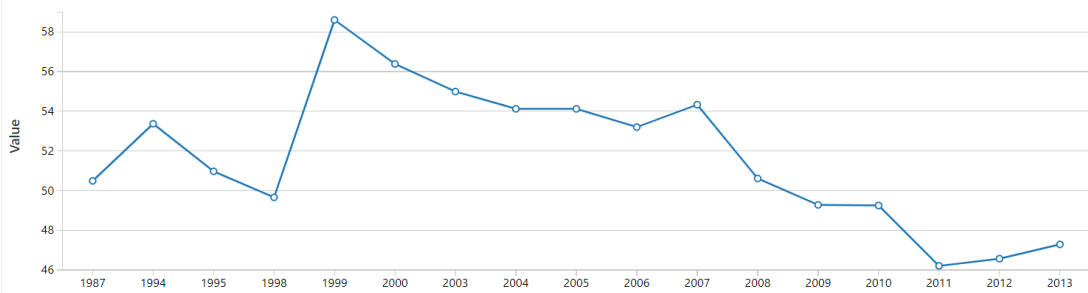▸ (1) Spark Jobs

GINI Index of Colombia

`SQL` ►▾ �🔢 ⌄ ━ ✕

```sql
1   %sql
2
3   select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "COL" order by Year asc;
```

▸ (1) Spark Jobs

Visualization ⌄ ＋

▸ (1) Spark Jobs

⤓ ✎ Edit Visualization   20 rows | 20.67 seconds runtime   Refreshed 33 days ago
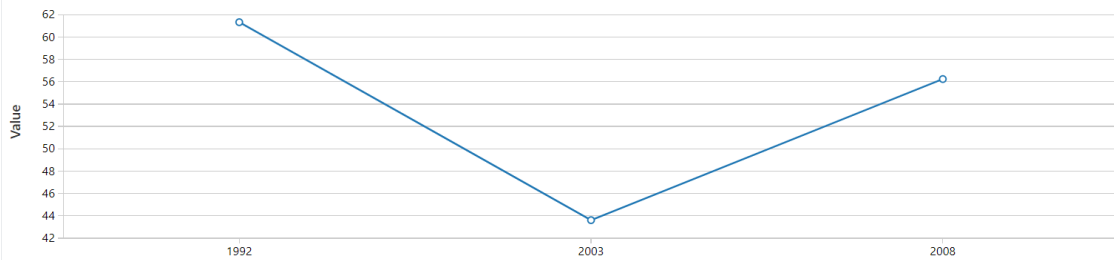
GINI Index of Rwanda

`SQL` ►▾ ⌄ ━ ✕

```sql
1   %sql
2
3   select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "RWA" order by Year asc;
```

▸ (1) Spark Jobs

Visualization ⌄ ＋

▸ (1) Spark Jobs

GINI Index of Russia

```
1   %sql
2
3   select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "RUS" order by Year asc;
```

▸ (1) Spark Jobs

Visualization ∨ +

▸ (1) Spark Jobs

Cmd 16

GINI Index of Ecuador

```
1   %sql
2
3   select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "ECU" order by Year asc;
```

▸ (1) Spark Jobs

Visualization ∨ +

▸ (1) Spark Jobs

GINI Index of Central African Republic

```
1  %sql
2
3  select Year,Value from Indicators where IndicatorCode  = "SI.POV.GINI" and CountryCode = "CAF" order by Year asc;
```

▸ (1) Spark Jobs

Visualization ∨ +
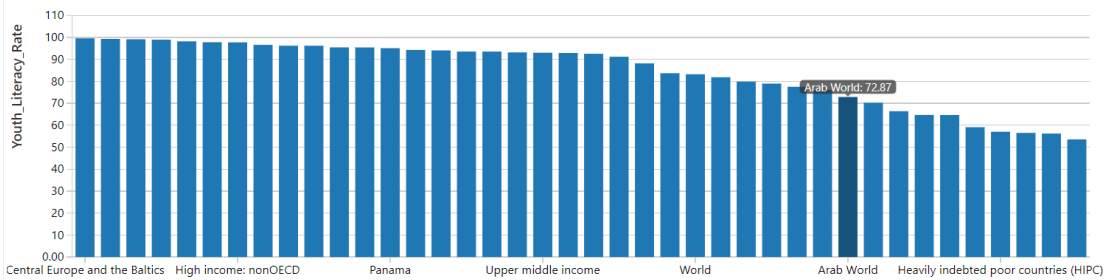
▸ (1) Spark Jobs
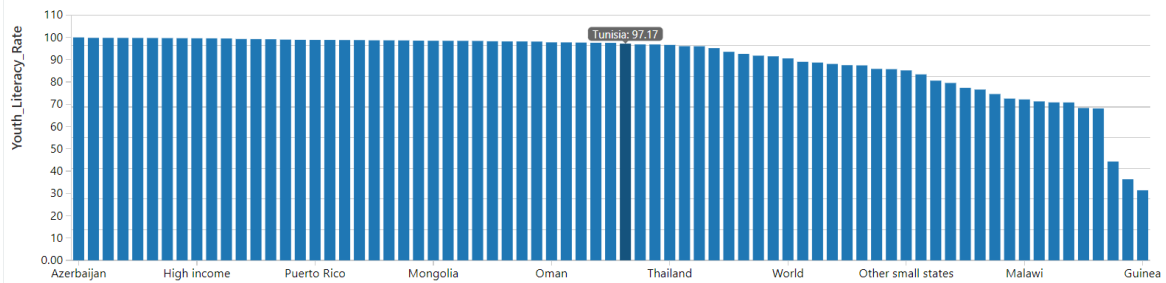
Barplot of Youth Literacy Rate in 1990

```
1  %sql
2
3  select value as Youth_Literacy_Rate,ShortName from Indicators A JOIN Country N ON A.CountryCode = N.CountryCode  where IndicatorCode  = "SE.ADT.1524.LT.ZS" and Year
   = 1990 order by Youth_Literacy_Rate Desc;
```

▸ (2) Spark Jobs

Visualization ∨ +

▸ (2) Spark Jobs

Cmd 19

Barplot of Youth Literacy Rate in 2010

SQL ▶▾ 📊 ⌄ — ✕

```
1  %sql
2
3  select value as Youth_Literacy_Rate,ShortName from Indicators A JOIN Country N ON A.CountryCode = N.CountryCode  where IndicatorCode  = "SE.ADT.1524.LT.ZS" and Year
   = 2010 order by Youth_Literacy_Rate Desc;
```

▸ (2) Spark Jobs

Visualization ⌄   +

▸ (2) Spark Jobs

Cmd 20

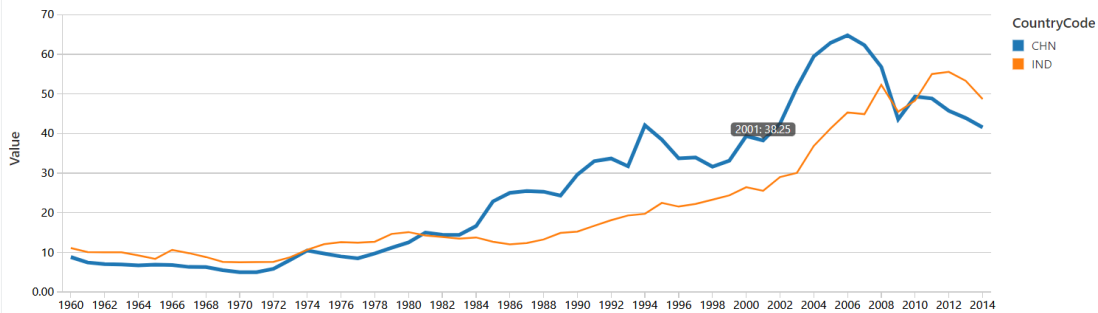## Trade as a percentage of GDP for China and India

SQL ▶▾ 📊 ⌄ — ✕

```
1  %sql
2
3  select Value,Year,CountryCode  from Indicators where IndicatorCode = "NE.TRD.GNFS.ZS" and CountryCode in ("IND", "CHN") order by Year;
```

▸ (1) Spark Jobs

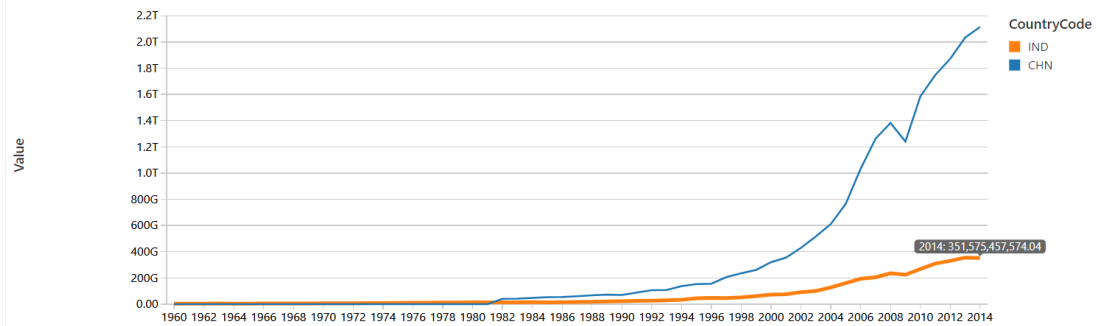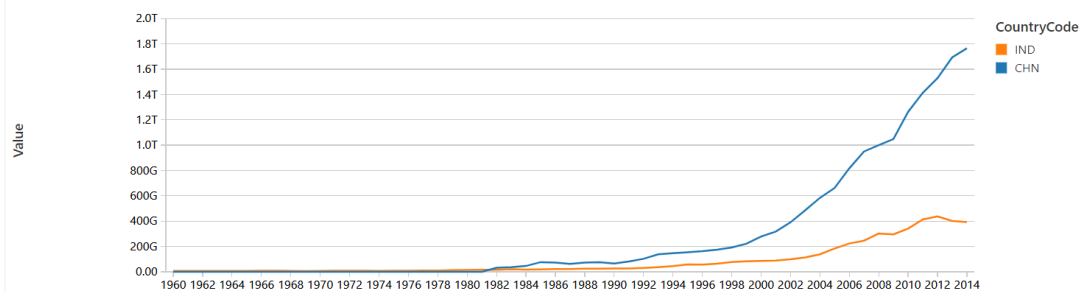Visualization ⌄   +

▸ (1) Spark Jobs

Exports of goods and services (constant 2005 US$)

```sql
1   %sql
2
3   select Value,Year,CountryCode  from Indicators where IndicatorCode = "NE.EXP.GNFS.KD" and CountryCode in ("IND", "CHN") order by Year;
```

▸ (1) Spark Jobs

Visualization ⌄   +

▸ (1) Spark Jobs

CountryCode
- IND
- CHN

Cmd 22

Import of goods and services (constant 2005 US$)

```sql
1   %sql
2
3   select Value,Year,CountryCode  from Indicators where IndicatorCode = "NE.IMP.GNFS.KD" and CountryCode in ("IND", "CHN") order by Year;
```

▸ (1) Spark Jobs

Visualization ⌄   +

▸ (1) Spark Jobs

CountryCode
- IND
- CHN

GDP per capita (adjusted by purchasing power parity)
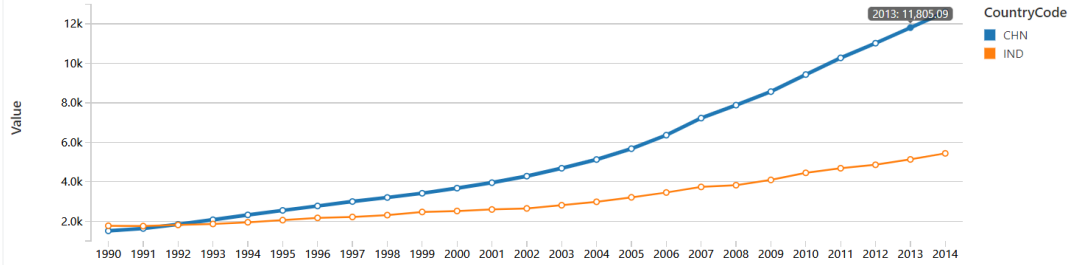
```
1  %sql
2
3  select Value,Year,CountryCode  from Indicators where IndicatorCode = "NY.GDP.PCAP.PP.KD" and CountryCode in ("IND", "CHN") order by Year;
```

▶ (1) Spark Jobs

Visualization ⌄   +

▶ (1) Spark Jobs

Poverty Alleviation
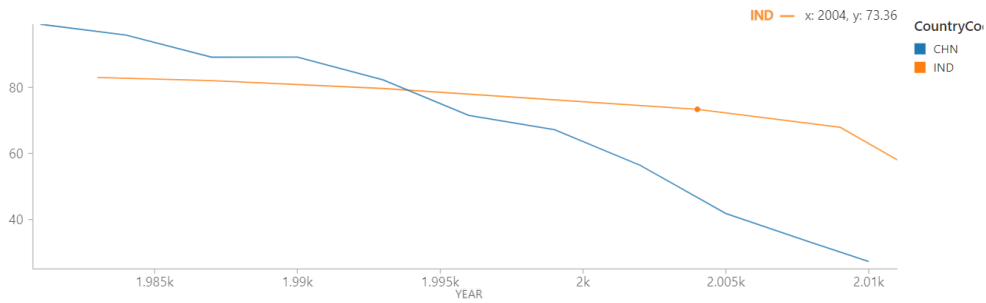
```
1  %sql
2
3  select Value,Year,CountryCode  from Indicators where IndicatorCode = "SI.POV.2DAY" and CountryCode in ("IND", "CHN") order by Year;
```

▶ (1) Spark Jobs

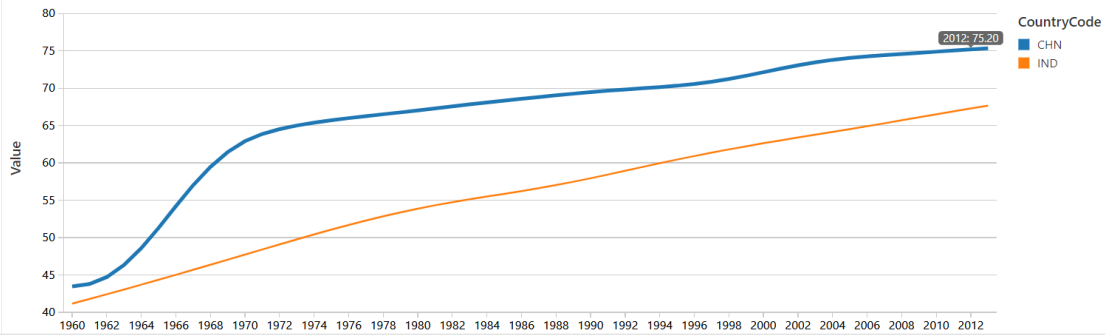Visualization ⌄   +

▶ (1) Spark Jobs

Life Expectancy at birth, total (years)

```
1   %sql
2
3   select Value,Year,CountryCode  from Indicators where IndicatorCode = "SP.DYN.LE00.IN" and CountryCode in ("IND", "CHN") order by Year;
```

▶ (1) Spark Jobs

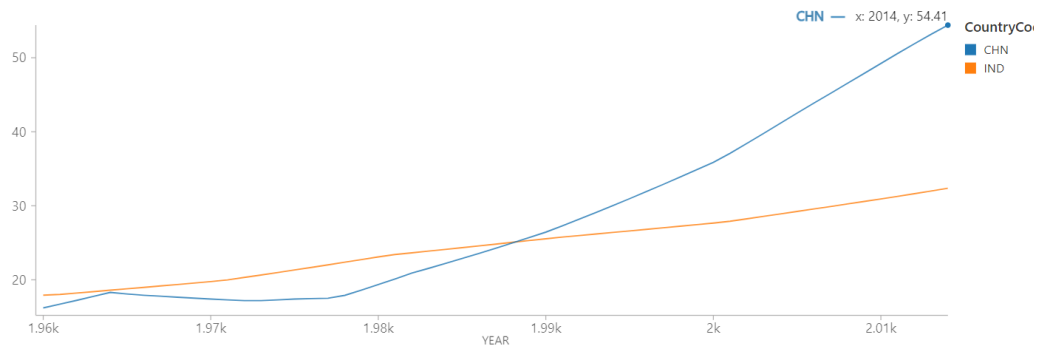**Visualization** ⌄ +

▶ (1) Spark Jobs

Urban Population growth

```
1   %sql
2
3   select Value,Year,CountryCode  from Indicators where IndicatorCode = "SP.URB.TOTL.IN.ZS" and CountryCode in ("IND", "CHN") order by Year;
```

▶ (1) Spark Jobs

Visualization ⌄ +

▶ (1) Spark Jobs

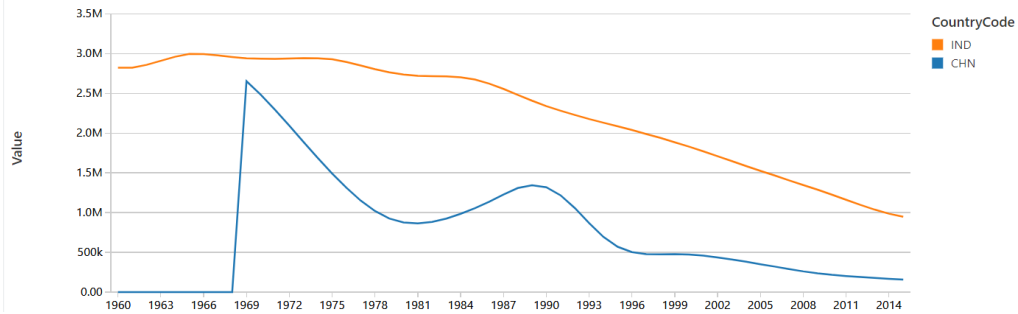## Infant Mortality - as a measure of health care

SQL

```sql
1   %sql
2
3   select Value,Year,CountryCode  from Indicators where IndicatorCode = "SH.DTH.IMRT" and CountryCode in ("IND", "CHN") order by Year;
```

▸ (1) Spark Jobs

Visualization ∨   +

▸ (1) Spark Jobs

CountryCode
- IND
- CHN

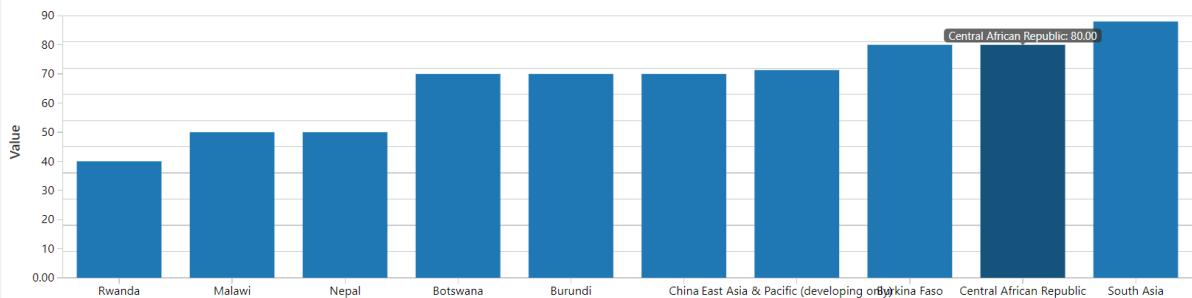## The 10 Countries with Lowest Average Income in 1962

SQL

```sql
1   %sql
2
3   select CountryName,Value from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and Year = 1962 order by Value asc limit 10;
```

▸ (1) Spark Jobs

Visualization ∨   +

▸ (1) Spark Jobs

Central African Republic: 80.00

Cmd 29

The 10 Countries with Lowest Average Income in 2014

```sql
1   %sql
2
3   select CountryName,Value from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and Year = 2014 order by Value asc limit 10;
```
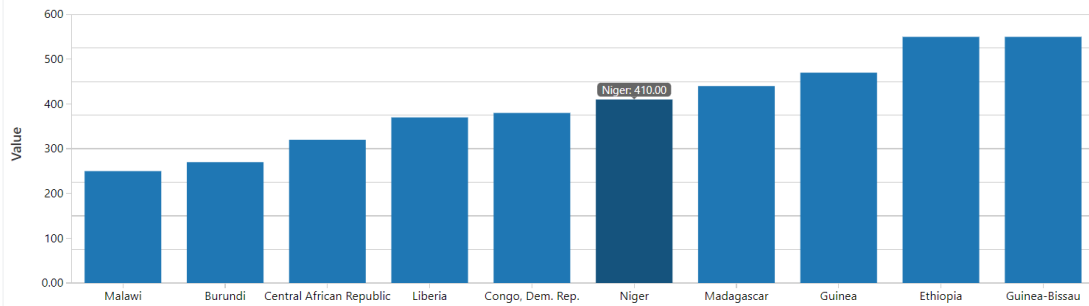
▶ (1) Spark Jobs

Visualization ⌄   +

▶ (1) Spark Jobs

Cmd 30

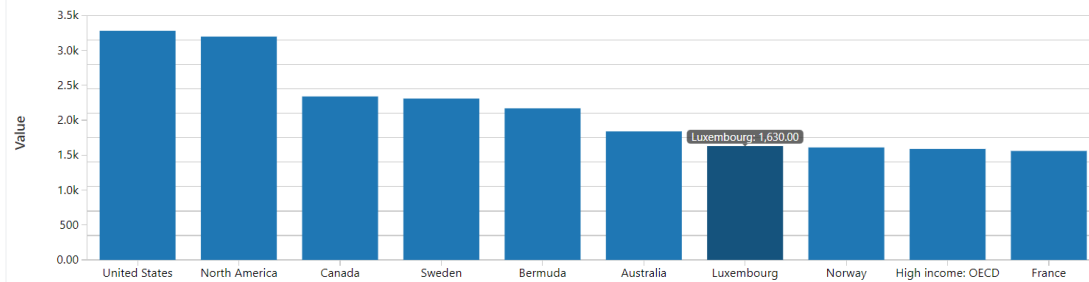## The 10 Countries with Highest Average Income in 1962

```sql
1   %sql
2
3   select CountryName,Value from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and Year = 1962 order by Value desc limit 10;
```

▶ (1) Spark Jobs

Visualization ⌄   +

▶ (1) Spark Jobs

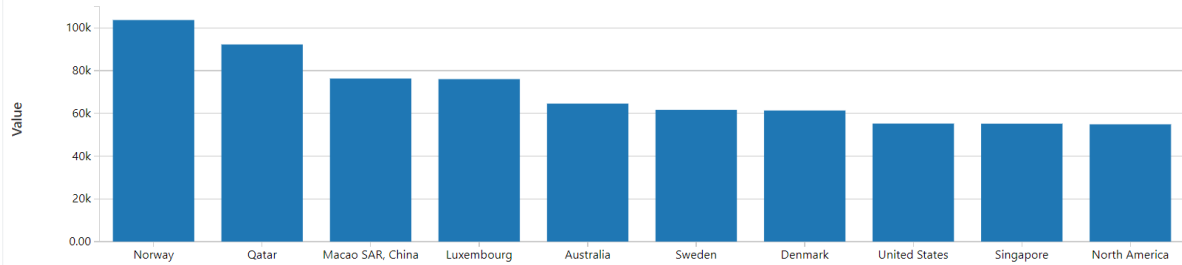## The 10 Countries with Highest Average Income in 2014

SQL ▶▼ ▢ ⌄ — ✕

```
1   %sql
2
3   select CountryName,Value from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and Year = 2014 order by Value desc limit 10;
```

▸ (1) Spark Jobs

Visualization ⌄ +

▸ (1) Spark Jobs

Average Income from 1960-2014 in Rich Countries

SQL ▶▼ ▢ ⌄ — ✕
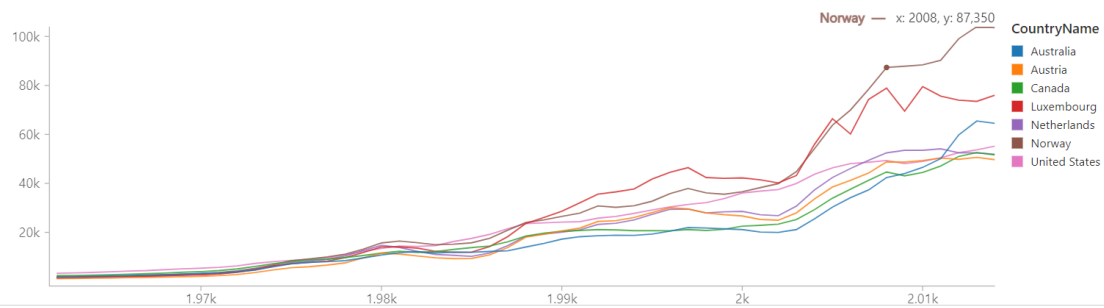
```
1   %sql
2
3   select CountryName,Value,Year from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and CountryName in("Australia","Austria","Canada", "Luxembourg",
    "Netherlands","Norway","United States")
4
```

▸ (3) Spark Jobs

Visualization ⌄ +

▸ (3) Spark Jobs

Average Income from 1960-2014 in Poor Countries

```sql
1  %sql
2
3  select CountryName,Value,Year from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and CountryName in("Burundi","Togo","Malawi","Central African Republic")
4
```
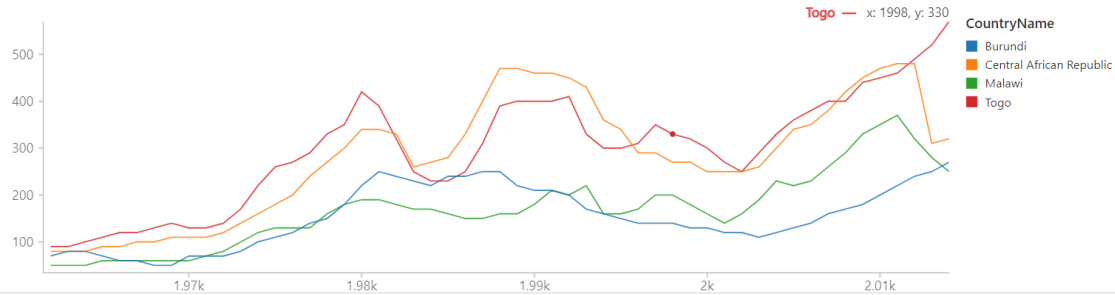
▸ (3) Spark Jobs

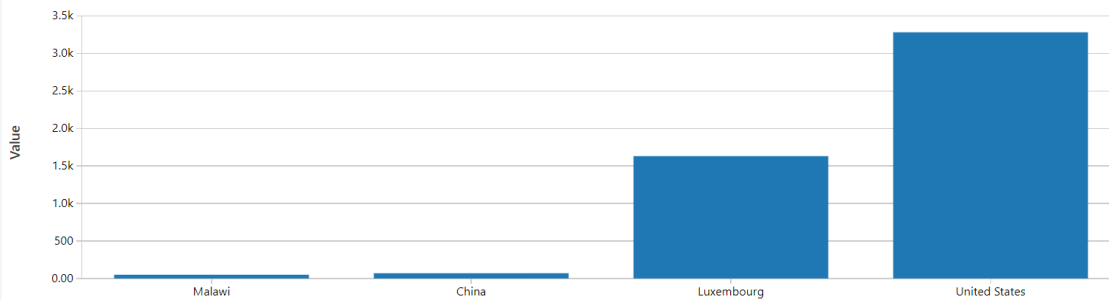**Visualization** ⌄   +

▸ (3) Spark Jobs

Average Income in 1962

```sql
1  %sql
2
3  select CountryName,Value,Year from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and Year = 1962 and CountryName in ("Malawi","China","Luxembourg","United States") order by Value asc;
```

▸ (1) Spark Jobs

Visualization ⌄   +

▸ (1) Spark Jobs

Average Income in 2014

`SQL` ▶▼ ☐ ⌄ ─ ✕

```sql
1  %sql
2
3  select CountryName,Value,Year from Indicators where IndicatorCode in ("NY.GNP.PCAP.CD") and Year = 2014 and CountryName in ("Malawi","China","Luxembourg","United States") order by Value asc;
```

▶ (1) Spark Jobs

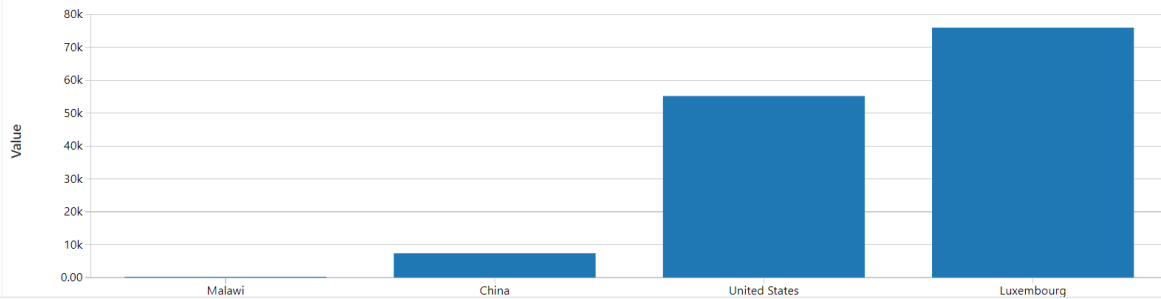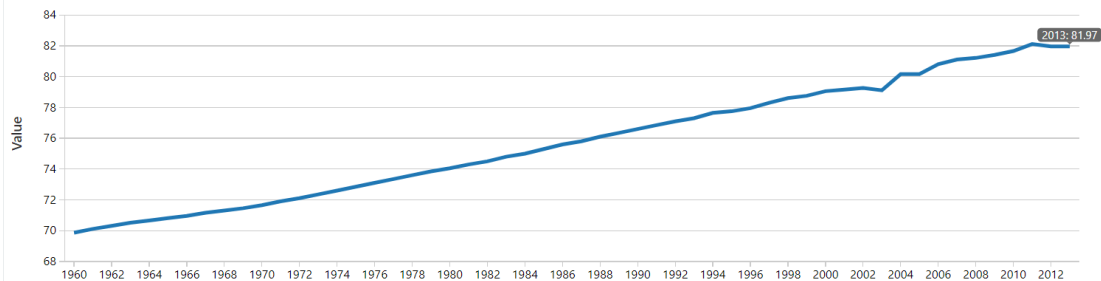**Visualization** ⌄   +

▶ (1) Spark Jobs

Life Expectancy in France 1960-2013

`SQL` ▶▼ ☐ ⌄ ─ ✕

```sql
1  %sql
2  SELECT year,Value
3      FROM Indicators
4      WHERE CountryName = "France"
5      AND IndicatorCode = "SP.DYN.LE00.IN";
```

▶ (3) Spark Jobs

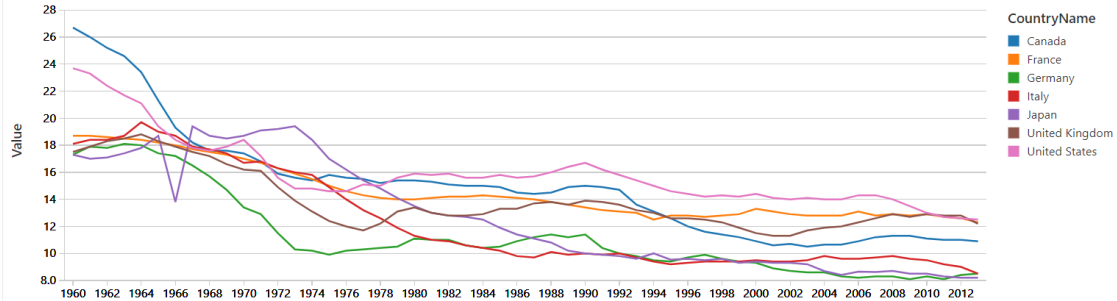**Visualization** ⌄   +

▶ (3) Spark Jobs

G-7 Country Birth Rates 1960-2013

```
1  %sql
2
3  SELECT CountryName,IndicatorName,IndicatorCode,Year,Value FROM Indicators WHERE IndicatorCode = 'SP.DYN.CBRT.IN' AND CountryName IN ('Canada','France','Germany',
   'United Kingdom', 'Italy', 'Japan','United States')
```

▸ (3) Spark Jobs

| Visualization ⌄ | + |

▸ (3) Spark Jobs

World Per Capita Income in 2013

```
1  %sql
2  SELECT countrycode,int(value) FROM Indicators WHERE Year = 2013 AND IndicatorCode = 'NY.ADJ.NNTY.PC.KD' order by value desc limit 250;
```
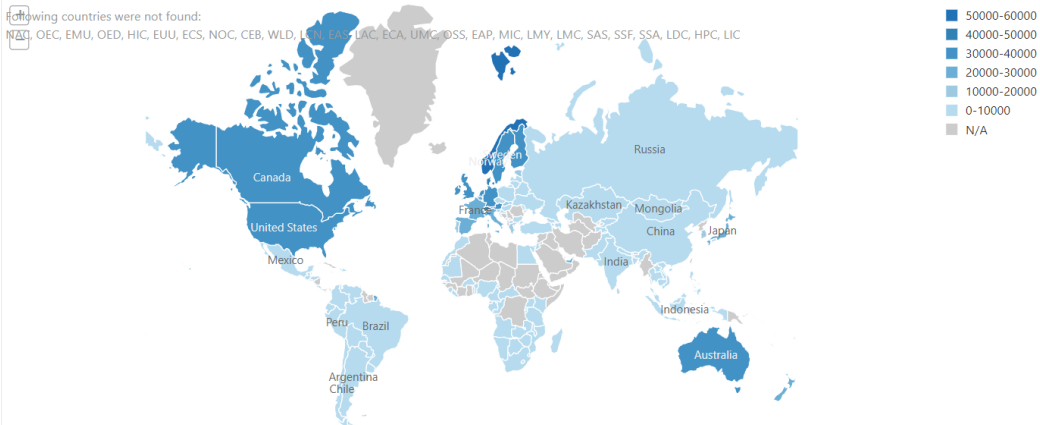
▸ (1) Spark Jobs

| Visualization ⌄ | + |

▾ (1) Spark Jobs

▸ Job 52   View (Stages: 1/1)

▼ (1) Spark Jobs
  ▸ Job 52   View (Stages: 1/1)

Following countries were not found:
NAC, OEC, EMU, OED, HIC, EUU, ECS, NOC, CEB, WLD, LDN, EAS, LAC, ECA, UMC, OSS, EAP, MIC, LMY, LMC, SAS, SSF, SSA, LDC, HPC, LIC

	50000-60000
	40000-50000
	30000-40000
	20000-30000
	10000-20000
	0-10000
	N/A

Canada

United States

Mexico

Peru   Brazil

Argentina
Chile

Sweden
Norway

France

Russia

Kazakhstan   Mongolia

China   Japan

India

Indonesia

Australia

Morocco: 1897

⤓   ✎ Edit Visualization   143 rows | 19.31 seconds runtime