

Assignment 4 Data

在课程网站上直接下载 countriesData.csv 文件。文件共包含 215 个国家, 2010-2014 年, 17 项条目的数据。每个国家的数据参考下图:

	2010	2011	2012	2013	2014
GDP growth (annual %)	11	9	8	8	7
GDP (current US\$)	6,039,658,508,486	7,492,432,097,810	8,461,623,162,714	9,490,602,600,148	10,360,105,247,908
Internet users (per 100 people)	34	38	42	46	49
Rural population	679,206,337	664,363,135	649,832,871	635,688,202	621,970,693
Population, female (% of total)	49	49	49	48	48
Population growth (annual %)	0	0	0	0	1
Population, total	1,337,705,000	1,344,130,000	1,350,695,000	1,357,380,000	1,364,270,000
GDP per capita growth (annual %)	10	9	7	7	7
GDP per capita (current US\$)	4,515	5,574	6,265	6,992	7,594
Labor force, total	781,054,640	790,183,028	795,863,202	801,790,592	..
Labor force, female (% of total labor force)	44	44	44	44	..
Armed forces personnel (% of total labor force)	0	0	0	0	..
Surface area (sq. km)	9,562,911	9,562,911	9,562,911	9,562,911	9,562,911
Population density (people per sq. km of land area)	142	143	144	145	145
Arable land (% of land area)	11	11	11	11	..
Unemployment, total (% of total labor force) (modeled ILO estimate)	4	4	5	5	..
Unemployment, female (% of female labor force) (modeled ILO estimate)	3	4	4	4	..

	2010	2011	2012	2013	2014
GDP 增长率 (年百分比)	11	9	8	8	7
GDP (现价美元)	6,039,658,508,486	7,492,432,097,810	8,461,623,162,714	9,490,602,600,148	10,360,105,247,908
互联网用户 (每 100 人)	34	38	42	46	49
农村人口	679,206,337	664,363,135	649,832,871	635,688,202	621,970,693
人口, 女性 (占总人口的百分比)	49	49	49	48	48
人口增长 (年度百分比)	0	0	0	0	1
人口, 总数	1,337,705,000	1,344,130,000	1,350,695,000	1,357,380,000	1,364,270,000
人均 GDP 增长 (年增长率)	10	9	7	7	7
人均 GDP (现价美元)	4,515	5,574	6,265	6,992	7,594
劳动力, 总数	781,054,640	790,183,028	795,863,202	801,790,592	..
劳动力, 女性 (占劳动力总数的百分比)	44	44	44	44	..
武装部队人员 (占劳动力总数的百分比)	0	0	0	0	..
国土面积 (平方千米)	9,562,911	9,562,911	9,562,911	9,562,911	9,562,911
人口密度 (每公里土地面积人数)	142	143	144	145	145
耕地 (占土地面积的百分比)	11	11	11	11	..
总失业人数 (占劳动力总数的比例) (模拟劳工组织估计)	4	4	5	5	..
女性失业人数 (占女性劳动力比例) (模拟劳工组织估计)	3	4	4	4	..

图 1 中国的数据内容

在 CSV 文件中,第一行为数据属性名称,包括 Country Name, Country Code, Series Name, Series Code, 2010 [YR2010], 2011[YR2011], 2012[YR2012], 2013[YR2013], 2014[YR2014].其中 Series Name 可能取值有 17 中,分别是 GDP growth (annual %),GDP (current US\$),Internet users (per 100 people), Rural population, Population, female (% of total), Population growth (annual %), Population, total, GDP per capita growth (annual %), GDP per capita (current US\$), Laborforce,total, Laborforce,female(%oftotallaborforce), Armedforcespersonnel (% of total labor force), Surface area (sq. km), Population density (people per sq. km of land area), Arableland(%oflandarea), Unemployment, total(%oftotallaborforce) (modeledILO estimate), Unemployment, female (% of female labor force) (modeled ILO estimate)。注意:csv 文件中,存在数据缺失现象。比如中国,劳动力总数在 2014 年是没有数据的。csv 文件中,使用“..”表示数据的缺失。建议大家开始设计前,先理解数据,选择某一些角度对数据进行分析。