Luke He (lh544)
Amanda Ong (ayo3)
Benjamin Steeper (bds238)

An Analysis of Urbanization and Its Effects on Education & Information Access

A.

Our data visualization focuses on visualizing effects of urbanization on information access through internet usage, educational attainment (primary school), government expenditures on education, & literacy rates. These factors were measured as a % of total population who used the internet, % of total population who graduated from primary school, % of total government expenditures, and % of total population who were literate, respectively. Urbanization was measured as % of total population who lived in urban cities.

Source of the Data and Processing It

Links to Our Datasets:

WDI Data: https://datacatalog.worldbank.org/dataset/world-development-indicators
Education Data: (https://datacatalog.worldbank.org/dataset/education-statistics

Our data comes from the World Bank's collection of publicly available data; specifically, its data on education statistics by country and world development indicators (WDI) by country (links provided above). These data sets were both incredibly large. Information was given both based on region (e.g. "Arab World" or "European countries") as well as by country, and provided a countless number of metrics over the years 1960 - 2013 (for WDI data) or 1970 - 2013 (for educational data). We decided, for consistency between the two data sets, to only look at data from 1970 to 2013. We decided on approximately eight to ten metrics that we judged to be the most indicative of a country's growth and circumstances, then reduced these chosen metrics to four based on the availability of the data for each of the years. For example, we first chose to look at the rate of educational attainment at the Bachelor/university level, but a majority of countries did not have this data available. We decided, in order to be comprehensive, to include every single country in our visualization; any metrics for which a country didn't have information available in the data set was not displayed in the visualization at all. We settled on the following four metrics, which we judged to be the most representative of a country's educational circumstances and had enough data available for a meaningful visualization:

1) Literacy rate
2) Percent of government expenditures spent towards education
3) Educational attainment (at the primary level)
4) Internet usage (as a percent of population that have access)

To extract the information we wanted from the dataset, we wrote a Python script (included the .zip file as "condense_data.py"). We reorganized the data to more easily work with it in d3 later. Each entry in the original data set provided the country, the indicator name, and the value of the indicator for every year from 1960 (or 1970, in the education statistics' case) to 2013. Our new data set organized the data such that every row represents a single year for a single country, and every column is the value of an indicator at that time, for that country. Although we settled on four metrics, we included all metrics we considered (and eventually eliminated) in order to be comprehensive.

To better visualize how we processed the data, below is a screenshot of the original and the processed data:

Original Data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Country N	Country C	Indicator Name	Indicator C	1960	1961	1962	1963	1964	1965	1966	1967	1968
188399	Germany	DEU	Population ages 45-49, male (% of male population)	SP.POP.45	6.221268	5.730016	5.131286	4.530183	4.084286	3.884635	3.88769	4.128677	4.52525
188400	Germany	DEU	Population ages 50-54, female (% of female population)	SP.POP.50	7.758546	7.752083	7.740672	7.681796	7.507266	7.18243	6.684688	6.018686	5.318915
188401	Germany	DEU	Population ages 50-54, male (% of male population)	SP.POP.50	6.966036	6.776258	6.597669	6.398475	6.129036	5.766779	5.28914	4.721008	4.157144
188402	Germany	DEU	Population ages 55-59, female (% of female population)	SP.POP.55	7.048414	7.125475	7.194898	7.252033	7.294359	7.318559	7.302186	7.300543	7.263249
188403	Germany	DEU	Population ages 55-59, male (% of male population)	SP.POP.55	6.836351	6.804995	6.70829	6.570179	6.424651	6.287132	6.109529	5.955952	5.790883
188404	Germany	DEU	Population ages 5-9, female (% of female population)	SP.POP.05	6.583963	6.738936	6.844496	6.901571	6.936779	6.990864	7.08838	7.21298	7.356186
188405	Germany	DEU	Population ages 5-9, male (% of male population)	SP.POP.05	7.920399	8.071155	8.16069	8.193271	8.202164	8.234945	8.331819	8.461064	8.611249
188406	Germany	DEU	Population ages 60-64, female (% of female population)	SP.POP.60	6.143094	6.227422	6.299407	6.366425	6.441585	6.530092	6.578027	6.637259	6.700761
188407	Germany	DEU	Population ages 60-64, male (% of male population)	SP.POP.60	5.246934	5.386905	5.540288	5.6887	5.805248	5.874078	5.84398	5.77397	5.680671
188408	Germany	DEU	Population ages 65 and above (% of total)	SP.POP.65	11.47399	11.68605	11.87978	12.06522	12.25776	12.46608	12.69781	12.93041	13.16109
188409	Germany	DEU	Population ages 65 and above, female	SP.POP.65	4992030	5152206	5312857	5474806	5634397	5804356	5960315	6095834	6234478
188410	Germany	DEU	Population ages 65 and above, female (% of total)	SP.POP.65	12.81617	13.14767	13.46692	13.78113	14.10034	14.43008	14.718	15.00351	15.29286
188411	Germany	DEU	Population ages 65 and above, male	SP.POP.65	3362742	3423156	3481663	3539799	3597855	3665339	3765686	3853519	3937628
188412	Germany	DEU	Population ages 65 and above, male (% of total)	SP.POP.65	9.930172	10.01203	10.07	10.11732	10.17521	10.25566	10.43024	10.60934	10.78003
188413	Germany	DEU	Population ages 65 and above, total	SP.POP.65	8354776	8574950	8794097	9014450	9232343	9469694	9726565	9950127	10172773
188414	Germany	DEU	Population ages 65-69, female (% of female population)	SP.POP.65	4.960434	5.050369	5.152493	5.263066	5.378163	5.494805	5.547052	5.603288	5.667672
188415	Germany	DEU	Population ages 65-69, male (% of male population)	SP.POP.65	3.752855	3.797354	3.866066	3.955163	4.065107	4.197232	4.309704	4.446348	4.587211
188416	Germany	DEU	Population ages 70-74, female (% of female population)	SP.POP.70	3.737681	3.803143	3.867234	3.939287	4.029905	4.142767	4.204021	4.284218	4.380729
188417	Germany	DEU	Population ages 70-74, male (% of male population)	SP.POP.70	2.895579	2.855673	2.802139	2.755066	2.738291	2.761478	2.787297	2.831046	2.892397
188418	Germany	DEU	Population ages 75-79, female (% of female population)	SP.POP.75	2.368416	2.430461	2.500398	2.577662	2.659599	2.746266	2.794878	2.843022	2.899996
188419	Germany	DEU	Population ages 75-79, male (% of male population)	SP.POP.75	1.892779	1.896243	1.898926	1.898088	1.888381	1.869367	1.842934	1.804668	1.769792
188420	Germany	DEU	Population ages 80 and above, female (% of female population)	SP.POP.80	1.74964	1.863702	1.946792	2.00111	2.03267	2.046237	2.172051	2.272979	2.344462
188421	Germany	DEU	Population ages 80 and above, male (% of male population)	SP.POP.80	1.388959	1.462758	1.502867	1.509007	1.483429	1.427582	1.490304	1.52728	1.530628
188422	Germany	DEU	Population density (people per sq. km of land area)	EN.POP.DNST		210.1728	212.0293	214.0015	215.7315	217.58	219.4034	220.4088	221.3912
188423	Germany	DEU	Population growth (annual %)	SP.POP.GR	0.782669	0.769854	0.879432	0.925875	0.805141	0.85319	0.834561	0.457209	0.444717
188424	Germany	DEU	Population in largest city	EN.URB.LC	3259833	3252346	3247311	3242284	3237257	3232253	3227249	3222253	3217257
188425	Germany	DEU	Population in the largest city (% of urban population)	EN.URB.LC	6.27154	6.182044	6.112887	6.041786	5.978633	5.913414	5.849986	5.809124	5.769258
188426	Germany	DEU	Population in urban agglomerations of more than 1 million	EN.URB.M	6919572	6980524	6996913	7013766	7031118	7048908	7067211	7086016	7105358

Processed Data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Country na	Country cc	Year	Urbanization	CO2 emissions	Edu attainment (Bachelor)	Edu attainment (primary)	Fertility rate	Govt exp (	Govt exp (	Literacy rate	Enrollment ratio (primary)	Enrollment ratio (tertiary)	Internet usage	
4033	Indonesia	IDN	1996	37.235	1.266992998			2.627	1.0774	7.49682		110.0135803	11.54181004	0.056624	
4034	Indonesia	IDN	1997	38.406	1.373878882			2.579	1.07121	7.71022		109.946312	13.38344002	0.19491	
4035	Indonesia	IDN	1998	39.593	1.041245921			2.544				110.7724533	13.35931015	0.255307	
4036	Indonesia	IDN	1999	40.792	1.159992484			2.523				111.3793564	14.74929047	0.444416	
4037	Indonesia	IDN	2000	42.002	1.245241613			2.512				108.7780228	14.87967968	0.925564	
4038	Indonesia	IDN	2001	42.782	1.374818345			2.509	2.45455	11.59389		109.7656097	14.18721008	2.018614	
4039	Indonesia	IDN	2002	43.566	1.410233797			2.511	2.64835	14.36905		109.9067078	14.81107044	2.134136	
4040	Indonesia	IDN	2003	44.353	1.436404464			2.513	3.22388	16.27818		109.4723434	16.02807999	2.38702	
4041	Indonesia	IDN	2004	45.145	1.509898209			2.515	2.74348	14.1703	90.38479	108.829361	16.61841011	2.600286	
4042	Indonesia	IDN	2005	45.937	1.508480596			2.514	2.87726	15.1488		107.7922287	17.25680923	3.602025	
4043	Indonesia	IDN	2006	46.732	1.50157677		80.19657898	2.51			91.98227	106.0453873	17.31292915	4.764813	
4044	Indonesia	IDN	2007	47.528	1.611855395		71.69862366	2.505	3.04494	14.94431		109.0557404	17.81677055	5.786275	
4045	Indonesia	IDN	2008	48.327	1.763895139		71.87893677	2.499	2.90113	13.67475	92.1923	107.5075073	20.70027924	7.917479	
4046	Indonesia	IDN	2009	49.124	1.865165403		72.85478973	2.492	3.52274	19.30844	92.5817	108.1407089	23.05732918	6.92	
4047	Indonesia	IDN	2010	49.924	1.767907879			2.483	2.81397	16.65418		108.656311	24.19967079	10.92	
4048	Indonesia	IDN	2011	50.712	2.456844744		74.43771362	2.471	3.19015	18.00613	92.81191	108.6352463	26.50420952	12.28	
4049	Indonesia	IDN	2012	51.488	2.559750233			2.455	3.40578	18.09042		108.691452	30.65633011	14.52	
4050	Indonesia	IDN	2013	52.252	1.945094474			2.436	3.35768	17.60393		106.3430328	31.28556061	14.94	
4051	Indonesia	IDN	2014	53.003	1.819363319	7.966499805	76.60041809	2.414	3.27861	17.67214	95.11622	105.7409592	31.10210037	17.14	
4052	Iran, Islam	IRN	1970	41.212	3.222030784			6.44							
4053	Iran, Islam	IRN	1971	42.11	3.473981386			6.359	2.77184			71.66033173	2.970149994		
4054	Iran, Islam	IRN	1972	43.014	3.528449642			6.292	2.76721			78.29000854	3.35371995		
4055	Iran, Islam	IRN	1973	43.92	4.193447792			6.243	2.8753			81.51309204	3.771029949		
4056	Iran, Islam	IRN	1974	44.832	4.527580658			6.214	2.68378			84.30847931	3.860759974		
4057	Iran, Islam	IRN	1975	45.747	4.257479632			6.212	2.97021			93.19994354	4.109889984		
4058	Iran, Islam	IRN	1976	46.666	4.642963903			6.238			36.51839828	99.03990173	4.493969917		

B.

Data to Visual Elements

We first pulled all of the data from our condendsed_data.csv file using d3.csv, creating a dictionary housing all the necessary data. We used d3's enter() method to append circles to an SVG in a g element. For our comprehensive scatterplot, we translated the x and y axis 100 pixels away from the edge of the SVG. Our axises are both percentages, so domain ranged from 0 to 100 for each. After peer review, we decided a good size for our circles was 5px, and colored them with gentle, bright colors to dovetail nicely with the optimistic story our plots illustrate.

C.

Comprehensive Scatterplot

Our comprehensive scatterplot is the overlaying of all four individual scatterplots below, with 0.5 opacity in order to more clearly view overlapping clusters. The plot presents the four metrics described above: literacy rate, percent of government expenditures spent towards education, educational attainment (at the primary level), and internet usage (as a percent of population that have access). There is also onmouseover functionality built into this scatterplot. Hovering over a single point will enlarge that point and display that country's data with the correct metric pertaining to that point bolded on the left of the screen, below the key.

By sliding through the years with the slider at the bottom of the screen, one can sense a general trend as points shuffle to the top right corner of the plot. This signifies that as time progresses, urbanization rates increase, and so do most other metrics related to education, government spending and people's overall access to information. This shows that as a countries' urbanization increases, the people of that country gain better access to basic information through their literacy ability and education (books & primary school), through government spending on education, and through their access of the internet. This plot is designed to get the general point across that urbanization improves the lives people in these areas all over the world.

Individual scatterplots

We isolate each metric in the individual scatterplots. This allows users to view each metric's story independent of other data, all in relation to urbanization to get a more in-depth view of the positive effects of urbanization over time. Literacy rate points start out low and begin to shift upwards after the early 2000s when high urbanization rates were introduced. Educational attainment looks almost identical when analyzes side by side to Literacy rates. We were most surprised by our government expenditures metric (it is for this reason we found it important to include this variable). Although there is an increase over time, expenditures stay relatively low as time and urbanization progresses in comparison to other metrics. This probably has to do with government expenditures being less variable as a percentage by nature. Still though, there is a minor uptrend and an heavier accumulation of data moving upwards on both axises over time. Internet usage has the highest, most obvious upwards trend correlating with urbanization over time. In 1989, there was no data on internet usage, and in a matter of decades usership and access shot up forming a perfect positive correlation scatterplot by the late 2000s.

All in all, observing these four metrics transform through time tells a very optimistic story pertaining to the holistic effects of urbanization in various countries around the world. With its rise we see record levels related to educational attainment and access to information.