Data Visualization in *Capital in the 21st Century*

Noah Wright

The University of Texas Inequality Project

Lyndon B. Johnson School of Public Affairs

The University of Texas at Austin

May 19, 2015

UTIP Working Paper #70

Abstract:

This paper examines how data visualization is used to supplement the arguments in Thomas Piketty's *Capital in the 21st Century*. Piketty shows a consistent pattern of modifying his visualizations to provide stronger support for his arguments than his data contain, particularly with respect to r>g, one of the central arguments of the book. This modification takes the form of using disproportionate axis units and the addition of estimated or speculative context. The effect is to change the fundamental shape of the data trends, which can be clearly seen when proportional axes are used and hypothetical context removed.

Comments welcome to noahwright@gmail.com .

The concept of inequality in the modern era is inseparable from the quantitative data used to measure it. Unlike historical perspectives of inequality, which relied on narrative accounts of poverty and wealth, modern definitions of inequality virtually always include some type of statistic such as a Gini coefficient, Theil statistic, or share of wealth or income owned by population quantiles. Tables of data, however, do not lend themselves to effective communication with a mass audience, making visualization a critical aspect of modern inequality literature.

In this paper I will analyze how Thomas Piketty uses data visualization in *Capital in the 21st Century*. The book contains a plethora of visualizations, including 96 graphs and 18 tables, which Piketty uses to reinforce the arguments being made by the text. I will examine the integrity of these visualizations and how they interact with the arguments of *Capital in the 21st Century*.

Edward Tufte devotes an entire chapter to graphical integrity in *The Visual Display of Quantitative Information*, a seminal text in the history of data visualization. Tufte identifies general principles of data visualization integrity; for the purposes of this analysis, I will focus on the two most relevant to *Capital in the 21st Century*: proportionality and context. Tufte defines data proportionality as follows: "the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented." (Tufte 56). Tufte proposes a method of measuring the distortion of this proportionality called the Lie Factor, in which the size of the effect shown in the graphic is divided by the size of the effect shown in data. For example:





This is the exact graphic Tufte uses to demonstrate a high lie factor. The change in the graphic is 5.3 inches – 0.6 inches, or 4.7 inches, indicating a 4.7/0.6 (783%) change. The change in the data, however, is 27.5-18, a 9.5/18 (53%) change. The total lie factor of this graphic is therefore 783/53, or 14.8. Lie factors are not an exact science, but they provide a rough conceptual framework for thinking about how one chooses to represent data.

In addition to distortion of proportionality, the removal or addition of context can alter one's perception to the same extent. For example, this historical chart of the Dow Jones Industrial Average makes the stock market looks like a safe long-term investment circa May of 1929:



Source:http://www.macrotrends.net/1319/dow-jones-100-year-historical-chart

The visualizations in *Capital in the 21st Century's* come in two forms: tables of numbers and timeseries line graphs. Almost every line graph is identical: the x-axis is years and the y-axis is an inequality measure or economic statistic. The following is the first graph that appears in the book, income inequality in the United States from 1910-2010. (Piketty, 24)



Piketty describes the graph in his text as follows: [italics are mine]

The US curve, shown in Figure I.1, indicates the share of the upper decile of the income hierarchy in US national income from 1910 to 2010. It is nothing more than an extension of the historical series Kuznets established for the period 1913–1948. The top decile claimed as much as 45–50 percent of national income in the 1910s–1920s before *dropping* to 30–35 percent by the end of the 1940s.

Inequality then *stabilized* at that level from 1950 to 1970. We subsequently see a *rapid rise* in inequality in the 1980s, until by 2000 we *have returned to a level* on the order of 45–50 percent of national income. *The magnitude of the change is impressive. It is natural to ask how far such a trend might continue.* (Piketty, 23)

In his description, Piketty tells the story of this graph for the reader. Inequality was high,

inequality the "dropped", "stabilized", and saw a "rapid rise" that "returned" it to its previous level. This is one of the major arguments of his book in miniature: that the declining inequality of Western Europe and the United States observed in the middle parts of the 20th century was temporary, and that the world is returning to a society defined by inherited wealth as in the 19th century.

In terms of proportionality, this visualization has 1 to 1 correspondence between the data depicted on both axes. One unit of the X-axis represents ten years and one unit of the y-axis represents 5% of national income share controlled by top decile.

In terms of context, the graph depicts the top decile share of national income, but the text does not specify whether this includes taxes and transfers, many of which are untaxed.¹ It does specify that the data comes from tax records, but fails to mention one of the fundamental problems of using tax records: that they only measure reported and taxable income. Both of these factors change the results of the visualization: taxes and transfers reduce inequality, while underreporting of income in response to high tax periods (like the 1940's-1960's), or exclusion thanks to loopholes, could artificially reduce inequality for the duration of the period.

These data choices shape the graph into something that is visually consistent with Piketty's narrative: we are returning to the old world. This becomes apparent when we shift the Y-axis:

¹ A virtually identical version of this graph appears in "Striking it Richer: The Evolution of Top Incomes in the United States," a 2013 paper by Piketty's collaborator Emmanuel Saez. In this paper, Saez specifies that the data is pre-taxes and transfers. (Saez, 8)



By definition, the top decile must control at least 10% of the wealth, so starting the Y-axis at zero as above is nonsense. Nevertheless, there is a point to the comparison. Could Piketty have described inequality as rapidly rising in the 1980's, if he had used the above visualization? His own visualization makes the change in U.S. inequality look larger, which aligns the visual with what Piketty emphasizes in his narrative of the world returning to how it used to be. This message appears numerous times throughout the book, with Piketty making frequent reference to the works of Jane Austen and Honoré de Balzac in tandem with his economic arguments. Throughout the book, Piketty alters context and proportionality in his visualizations to fit them around this narrative. Of the 92 time-series graphs in the book, 25 have non-proportional X-axes.² The first of these to appear is Figures 1.1, seen on the next page: (Piketty, 60)

² See appendix A for a full list.



One unit on the x-axis represents 120 years, then 50 years, then 63 years, then 37 years, then 20 years, then 20 years again, then 22 years. Thus almost every unit represents a different amount of time. The inconsistency makes Tufte's "lie factor" difficult to calculate, but the maximum disproportion is 6-fold: 120 years vs. 20 years. When the graph is displayed with a proportional x-axis, it looks as follows:³



3 This resizing is drawn from Piketty's data, which only includes the figures for the years in the original graph (1700, 1820, 1913, etc.). A year-by-year dataset would yield a much smoother graph, see Appendix B for technical details regarding the visualizations used in this paper.

The effect of the disproportion is to emphasize the 20th century. Less than one third of the timeframe is the 20th century, yet this century occupies more than half of the figure. This emphasis also has the effect of visually depicting Piketty's narrative of a return to what used to be, as did figure I.1. Both figures depict a beginning state of affairs, a change in that state of affairs in the middle of the graph, and at the end a gradual return.

This specific proportionality is used as the x-axis labeling convention for figures 1.2, 1.3, 2.1, 2.3, 2.6, and all six of the figures in chapter 3. Sometimes this proportionality does not change the fundamental shape of the data trend, as seen in figure 2.1 on the next page. In these cases, the axis is likely chosen to make comparisons easier. In other cases, as seen in figure 2.3 on page 11, this proportionality reshapes the graph into a curve that depicts a status quo, an alteration, and a return to that status quo along the lines of what is depicted above.

Piketty's Graph: (Piketty, 75)



A proportional X-axis does not reshape the trend:



Figure 2.1. The growth of world population 1700-2012

World population rose from 600 million inhabitants in 1700 to 7 billion in 2012. Sources ans series: see piketty.pse.ens.fr/capital21c.

Piketty's graph: (Piketty, 97)



A proportional x-axis makes the trend look more flat:



In all of these cases, regardless of the change in shape, the years chosen from the 20th century stand out in these figures. The specific years were likely chosen due to the large role they play in the narrative of *Capital in the 21st Century*: 1913 (just prior to World War 1), 1950 (just after World War 2, when economic recovery is in full swing), 1970 (when that economic recovery begins to stall), and 1990 (when inequality is widening once more). These years continue to be emphasized even in much longer timeframes than the graphs above, leading to severely disproportionate visualizations. The first of these is a graph of the world population growth rate over time, seen below. (Piketty, 80)



The growth rate of world population was above 1% per year from 1950 to 2012 and should return toward 0% by the end of the 21st century. Sources and series: see piketly.pse.ens.fr/capital21c.

The proportionality of the x-axis does not come close to matching the scale of the data. As with the above graphs, almost every unit represents a different amount of time, but this time the highest degree of disproportion is over 50-fold: 1000 years vs. 18 years, both represented by one unit. If each unit on the X-axis had the same value, the graph would look as follows:



The growth rate of world population was above 1% per year from 1950 to 2012 and should return toward 0% by the end of the 21st century. Sources and series: see piketty.pse.ens.fr/capital21c.

Piketty's description of the data reveals the story he tells with it: [italics mine]

According to official forecasts, progress toward the demographic transition at the global level should now accelerate, leading to *eventual stabilization* of the planet's population. According to a UN forecast, the demographic growth rate should *fall* to 0.4 percent by the 2030s and *settle* around 0.1 percent in the 2070s. If this forecast is correct, the world will *return* to the very low-growth regime of the years before 1700. The global demographic growth rate would then have followed a *gigantic bell curve* in the period

1700–2100, with a *spectacular peak* of close to 2 percent in the period 1950–1990. (Piketty, 79, italics mine)

Piketty could not describe the proportional graph using the above language. While the

proportional graph does contain a spectacular peak, it does not resemble Piketty's sloping bell curve. This

graph shows the 20th century as an extremely brief, severe aberration, made all the more so by the

inclusion of long-term historical estimates and speculative projections. Piketty himself even

acknowledges that for 0 to 1700, the majority of this timescale:

-the precision of such estimates is illusory. We actually possess very little information about the growth of the world's population between 0 and 1700 and even less about output per head. Nevertheless, no matter how much uncertainty there is about the exact figures (which are not very important in any case), there is no doubt whatsoever that the pace of growth was quite slow from antiquity to the Industrial Revolution, certainly no more than 0.1–0.2 percent per year.

These historical estimates are "not very important", yet they are included in the visualization. The reason for doing so becomes clear when we remove these historical estimates and future projections in addition to rescaling the x-axis:



The growth rate of world population was above 1% per year from 1950 to 2012 and should return toward 0% by the end of the 21st century. Sources and series: see piketty.pse.ens.fr/capital21c.

This visualization does not look like a return to anything; the bell curve has disappeared. In addition to a disproportionate x-axis, Piketty has shaped the visualization by altering context. Rather than removing context, however, as I did in the stock market graph at the beginning of the paper, Piketty has added context to create the desired look for his visualization. For the above population graph, though, this additional data is questionable but not indefensible. The historical population growth rate has been very low from what we've been able to discern (if it hadn't been, there would be a great deal more people around now). Likewise, the observed population growth rate has been declining, and many if not most demographers and forecasters predict that it will continue to do so.

However, this addition of historical estimates and future projections becomes much more problematic in the next two graphs starting at antiquity, which depict growth rates of per-capita output and total output. Unlike population growth, the future projections in these are far less plausible. Piketty's own data does not show per-capita output declining in the real world, which means the projections of figure 2.4 showing this decline are entirely speculative. This alteration becomes much more apparent when Piketty's graphs as he presents them are compared with how they appear after the removal of projections, estimates, and disproportionate X-axes (as seen on the next page).

As with population, Piketty acknowledges the hypothetical nature of his projections in the text while simultaneously describing the phenomenon as a bell curve:

I have already conceded that these "median" forecasts are highly hypothetical. The key point is that regardless of the exact dates and growth rates (details that are obviously important), the *two bell curves of global growth* are in large part already determined. (Piketty, 102, italics mine)

Yet there is nothing resembling a bell curve once the above adjustments are made.

Piketty's Graph: (Piketty, 100)



Proportional X-axis, historical estimates and future projections removed:



Figure 2.4. The growth rate of world per capita output since Antiquity until 2100

The growth rate of per capita output surpassed 2% from 1950 to 2012. If the convergence process goes on, it will surpass 2,5% from 2012 to 2050, and then will drop below 1,5%. Sources and series : see piketty.pse.ens.fr/capital21c.





Proportional X-axis, historical estimates and future projections removed:



Figure 2.5. The growth rate of world output from Antiquity until 2100

The growth rate of world output surpassed 4% from 1950 to 1990. If the convergence process goes on it will drop below 2% by 2050. Sources and series: see piketty.pse.ens.fr/capital21c.

The chapter 2 time-series graphs seen above play an important role in laying the groundwork for the graphs in chapter 10, which depict one of the central arguments of *Capital in the 21st Century:* that when the rate of return on capital exceeds the general growth rate, inequality tends to increase. The first of these, figure 10.9, essentially replicates figure 2.5 in its depiction of the growth rate of world output while adding the trend-line for rate of return on capital:⁴ (Piketty 354)



Figure 10.9. Rate of return vs. growth rate at the world level,

Note that unlike in chapter 2, the hypothetical projections are not labeled as such. Piketty builds on this visualization by adding in the all of factors that reduced "r" in the 20th century to create Figure 10.10. (Piketty 356) Here the 20th century is presented as a clear outlier during which the growth rate of the economy exceeded the rate of return on capital. Contrast this with how the figure appears with a proportionate x-axis when historical estimates and future predictions are removed:

⁴ Many critiques have already been written about Piketty's notion of "rate of return on capital," see Acemoglu and Robinson (2015), Galbraith (2014), and McCloskey (2014). For this reason, I will be focusing solely on visualization in this paper for the sake of brevity.

Piketty's Graph: (Piketty 356)



Proportional X-axis, historical estimates and future projections removed:



Figure 10.10. After tax rate of return vs. growth rate at the world level, from Antiquity until 2100

The rescaled visualization does show the unique character of the 20th century, but gives no reason to believe that this trend won't continue into the future. This particular point is so important to Piketty that on the page following figure 10.10 is an almost-identical graph designed to even further emphasize the idea that the world is returning to its pre-20th century state: (Piketty, 357)



Figure 10.11. After tax rate of return vs. growth rate at the world level,

Figure 10.11 collapses the entire 20th century into a single point and adds another hundred years of purely speculative projection that Piketty himself acknowledges as a "hypothesis". (Piketty 357) Piketty justifies this decision as follows:

> To bring this possible evolution out even more clearly, I have combined in Figure 10.11 the two subperiods 1913–1950 and 1950–2012 into a single average for the century 1913–2012, the unprecedented era during which the net rate of return on capital was less than the growth rate. . .Figure 10.11 at least brings out the unprecedented-and possibly unique—character of the twentieth century in regard to the relation between r and g. (Piketty, 357)

These visualizations would not be nearly as problematic were it not for the fact that Piketty uses

them to make claims like the following:

As Figure 10.9 shows, the pure rate of return on capital—generally 4–5 percent—has *throughout history always* been distinctly greater than the global growth rate, but the gap between the two shrank significantly during the twentieth century, especially in the second half of the century, when the global economy grew at a rate of 3.5–4 percent a year. In all likelihood, *the gap will widen again* in the twenty-first century as growth (especially demographic growth) slows. (Piketty, 355, italics mine)

There is a circular logic at work here, as Figure 10.9 has been designed to show the conclusions being drawn from it. But without his reshaped x-axes, historical estimates, and future projections, Piketty cannot make the case that his central argument holds true "throughout history always", a claim restated in the ultimate conclusion of this series of graphs:

To sum up: the inequality r > g has clearly been true throughout most of human history, right up to the eve of World War I, and it will probably be true again in the twenty-first century. (Piketty, 358)

Prior to these graphs, Piketty refers to r > g as the "fundamental force for divergence" (Piketty 25), and goes on to refer to other components of his theory as "fundamental laws" (Piketty 52). A "fundamental law" is one of the strongest claims one can make in social sciences, as a fundamental law must always be true in every conceivable circumstance. If r > g is fundamental, it must be true throughout history and forever into the future. The strength of this claim necessitates these inherently problematic depictions of historical and projected future capital trends. Piketty could have avoided the need for these problematic depictions by weakening his assumptions and/or limiting himself to non-speculative data. Had he done this, however, he could never have justified one of the central claims of his book: that the world is returning to what it was in the past. What the data actually shows is that we do not know where the world is heading.

	Disproportionate	Pre-1700 or future		Disproportionate	Pre-1700 or future
Figure	X-axis?	estimates?	Figure	X-axis?	estimates?
0.1			8.2		
0.2			8.5		
1.1	Yes		8.6		
1.2	Yes		8.7		
1.3	Yes		8.8		
1.4			9.1		
1.5			9.2		
2.1	Yes		9.3		
2.2	Yes	Yes	9.4		
2.3	Yes		9.5		
2.4	Yes	Yes	9.6		
2.5	Yes	Yes	9.7		
2.6	Yes		9.8		
3.1	Yes		9.9		
3.2	Yes		10.1		
3.3	Yes		10.2		
3.4	Yes		10.3		
3.5	Yes		10.4		
3.6	Yes		10.5		
41	Yes		10.6		
4.2	Yes		10.7		
4 3	Yes		10.8		
4.4	105		10.0	Yes	Yes
4.5			10.10	Yes	Yes
4.6	Ves		10.10	Ves	Ves
4 7	Ves		11 1	105	105
4.8	Ves		11.1		Ves
49	105		11.2		Ves
4 10	Ves		11.5		105
5.1	105		11.1		
5.1			11.5		Ves
53			11.0		Ves
5.5			11.7		105
5.5			11.0		Vec
5.5			11.9		Ves
5.0			11.10		Ves
5.8		Vec	11.11		103
6.1		103	12.1		
6.2			12.1		
6.2			12.2		
6.4			12.5		Vac
6.5			12.4		Vec
0.3			12.5		res
0.0			12.0		
0./					
0.8					
8.1			14.2		
*Figures	4.11, 8.3, 8.4, 8.9, at	nd 8.10 are not time-s	eries and therefo	re not included abov	e.

Appendix A- Index of visualization characteristics in *Capital in the 21st Century*

The visualizations in this paper are drawn from the data made available by Thomas Piketty on his website for *Capital in the 21st Century*, <u>http://piketty.pse.ens.fr/en/capital21c2</u>. This data contains, for the most part, only what is presented in Piketty's figures. For example, Figure 2.2 appears in Piketty's spreadsheet as follows:

	World population (growth rate)	Europe	América	Africa	Asia
0-1000	0.0%	0.0%	0.1%	0.1%	0.0%
1000-1500	0.1%	0.2%	0.1%	0.1%	0.1%
1500-1700	0.2%	0.2%	-0.2%	0.1%	0.2%
1700-1820	0.5%	0.5%	0.7%	0.2%	0.5%
1820-1913	0.6%	0.8%	1.9%	0.6%	0.4%
1913-1950	0.9%	0.4%	1.6%	1.6%	0.9%
1950-1970	1.9%	0.9%	2.2%	2.4%	2.1%
1970-1990	1.8%	0.5%	1.7%	2.8%	2.0%
1990-2012	1.3%	0.1%	1.3%	2.4%	1.3%
2012-2030	0.9%	0.0%	0.8%	2.1%	0.8%
2030-2050	0.6%	-0.2%	0.4%	1.7%	0.3%
2050-2070	0.3%	-0.2%	0.1%	1.3%	-0.1%
2070-2100	0.1%	-0.1%	-0.1%	1.2%	-0.4%

(Sheet: Chapter2TablesFigures.xls, Tab: TS2.2)

There is additional data by region, but no additional data by year. In other words, the world population growth rate at 1760 or 1930 cannot be found in these tables, only the averaged values seen above. As a result, I produced my rescaled visualizations by reformatting the data as follows:

	World population				
	(growth rate)	Europe	América	Africa	Asia
1800	0.5%	0.5%	0.7%	0.2%	0.5%
1810	0.5%	0.5%	0.7%	0.2%	0.5%
1820	0.5%	0.8%	1.9%	0.6%	0.4%
1830	0.6%	0.8%	1.9%	0.6%	0.4%
1840	0.6%	0.8%	1.9%	0.6%	0.4%
1850	0.6%	0.8%	1.9%	0.6%	0.4%
1860	0.6%	0.8%	1.9%	0.6%	0.4%
1870	0.6%	0.8%	1.9%	0.6%	0.4%
1880	0.6%	0.8%	1.9%	0.6%	0.4%
1890	0.6%	0.8%	1.9%	0.6%	0.4%
1900	0.6%	0.8%	1.9%	0.6%	0.4%
1910	0.6%	0.4%	1.6%	1.6%	0.9%
1920	0.9%	0.4%	1.6%	1.6%	0.9%
1930	0.9%	0.4%	1.6%	1.6%	0.9%
1940	0.9%	0.4%	1.6%	1.6%	0.9%
1950	0.9%	0.9%	2.2%	2.4%	2.1%

-etc. You will notice that I use 1910 above when Piketty's data lists 1913. This is because of the way Excel automatically formats axis coordinates to be equidistant from one another regardless of scale. The only way around this would be to create a listing for every single year, but I chose decades because it is what Piketty does for most of his data points in these graphs. I have also used 2010 instead of 2012 and 2013 for the same reason.

Another problem I faced came from the fact that because these are annualized averages, multiple values could be chosen for the same year. To use the above table as an example again (I've copied the relevant section below), what should be used as the growth rate for 1970: 1.9%, 1.8%, or 1.85%?

Year	World population (growth rate)		
1950-1970	1.9%		
1970-1990	1.8%		

I can format it one of three ways:

Year	World population (growth rate)	Year	World population (growth rate)	Year		World population (growth rate)
1960	1.9%	1960	1.9%		1960	1.9%
1970	1.9%	1970	1.8%		1970	1.85%
1980	1.8%	1980	1.8%		1980	1.8%

-each of which produces a slightly different visual:



For the graphs I produced in this paper derived from averages, I went with the option on the left, using the first listed value for a given decade. Using the middle option would have created unevenly sized categories: the final 20 years would have had three entries as opposed to two. The option on the right I ruled out as extrapolation.

Works Cited

- Acemoglu, Daron & Robinson, James A. "The Rise and Decline of General Laws of Capitalism," Journal of Economic Perspectives, American Economic Association, vol. 29(1), pages 3-28, Winter 2015.
- Galbraith, James. "Kapital for the Twenty-First Century?" Dissent, Spring 2014. http://www.dissentmagazine.org/article/kapital-for-the-twenty-first-century
- McCloskey, Deirdre. "Measured, Unmeasured, Mismeasured, and Unjustified Pessimism: A Review Essay of Thomas Piketty's Capital in the Twenty-First Century," Erasmus Journal for Philosophy and Economics, Erasmus Institute for Philosophy and Economics, vol. 7(2), Autumn 2014.
- Piketty, Thomas & Goldhammer, Arthur. *Capital in the Twenty-First Century*. Cambridge, MA: The Belknap Press of Harvard University Press. 2015.
- Saez, Emmanuel. "Striking it Richer: The Evolution of Top Incomes in the United States," Berkeley, CA: UC Berkeley. 2014. < http://eml.berkeley.edu/~saez/saez-UStopincomes-2012.pdf>

Tufte, Edward. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press. 1983.