

SOC 3811/5811:
BASIC SOCIAL STATISTICS

Percentiles and Standardized Scores

More on Describing Distributions

Two more tools for describing continuous distributions

Percentiles

“Below what point in the distribution do 75% of the cases fall?”

Standardized (or “Z”) scores

“How many standard deviations from the mean does a particular observation fall?”

Percentiles

Percentile

The value “below which a given percentage of the observations in a distribution falls”

Why would we want to know the point on a distribution below which a given percentage of the cases fall?

Example: “Let’s give a tax break to people at or below the 80th percentile of the income distribution”

Example: “Let’s give a scholarship to high school students above the 95th percentile of the GPA distribution”

Percentiles

We saw something like this earlier when we talked about cumulative percentages

Values		Cumulative	Cumulative
Values	Frequency	Frequency	Percentage
0	5	5	16.7%
1	10	15	50.0%
2	10	25	83.3%
3	5	30	100.0%

Here it's pretty straightforward to figure out where the X^{th} percentile is; we simply find the lowest value at which the cumulative percentage equals or exceeds $X\%$

Percentiles

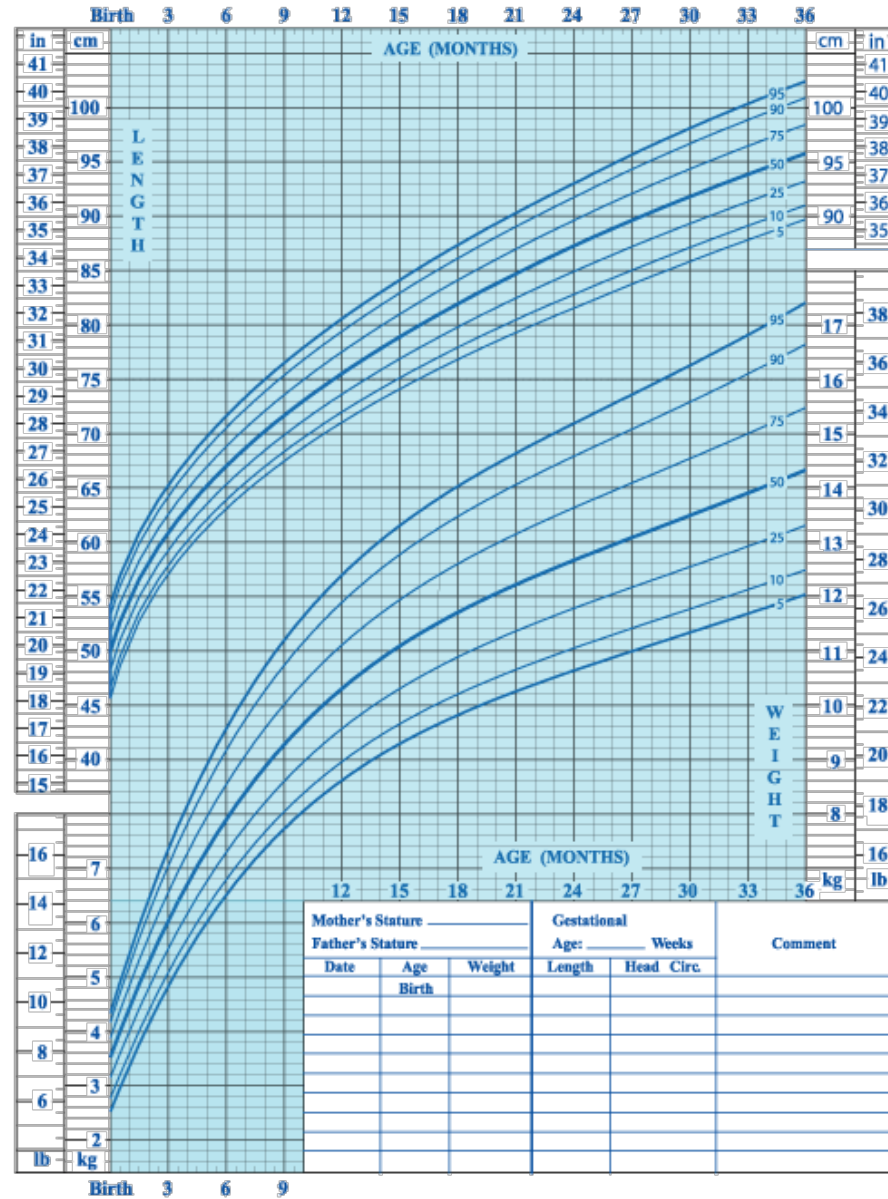
We saw something like this earlier when we talked about cumulative percentages

Values		Cumulative	Cumulative
Values	Frequency	Frequency	Percentage
0	5	5	16.7%
1	10	15	50.0%
2	10	25	83.3%
3	5	30	100.0%

What is the 40th percentile of this distribution? 40% of the cases fall at or below the value 1. What is the 75th percentile? 75% of the vases fall at or below 2

Birth to 36 months: Boys
Length-for-age and Weight-for-age percentiles

NAME _____



SOURCE : Developed by the National Center for Health Statistics. In collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). <http://www.cdc.gov/growthcharts>

NAME	SEX	DATE OF BIRTH	SOCIAL SECURITY NO	REPORT DATE
DAWN	E F			03/23/85 07079-R
TELEPHONE NUMBER	STATE AND COUNTY OF LEGAL RESIDENCE			REGISTRATION NUMBER
	34013 NJ ESSEX			2327159
ETHNIC GROUP	ENGLISH BEST LANGUAGE	U.S. CITIZEN	VETERAN	PARENTS' ADDRESS DIFFERENT
WHITE	YES	YES	NO	NO
				DATE OF SDQ
				JAN 85



ADMISSIONS
TESTING
PROGRAM
THE
COLLEGE
BOARD

HIGH SCHOOL INFORMATION			
H S CODE	H S NAME AND ADDRESS		
310750	COLUMBIA HIGH SCHOOL		
TYPE OF H S			
PUBLIC	MAPLEWOOD NJ 07040		
CLASS SIZE	SELF-REPORTED CLASS RANK	H S PROGRAM	H S GRADUATION DATE
250-499	2ND 10TH	GENERAL	JUN 85

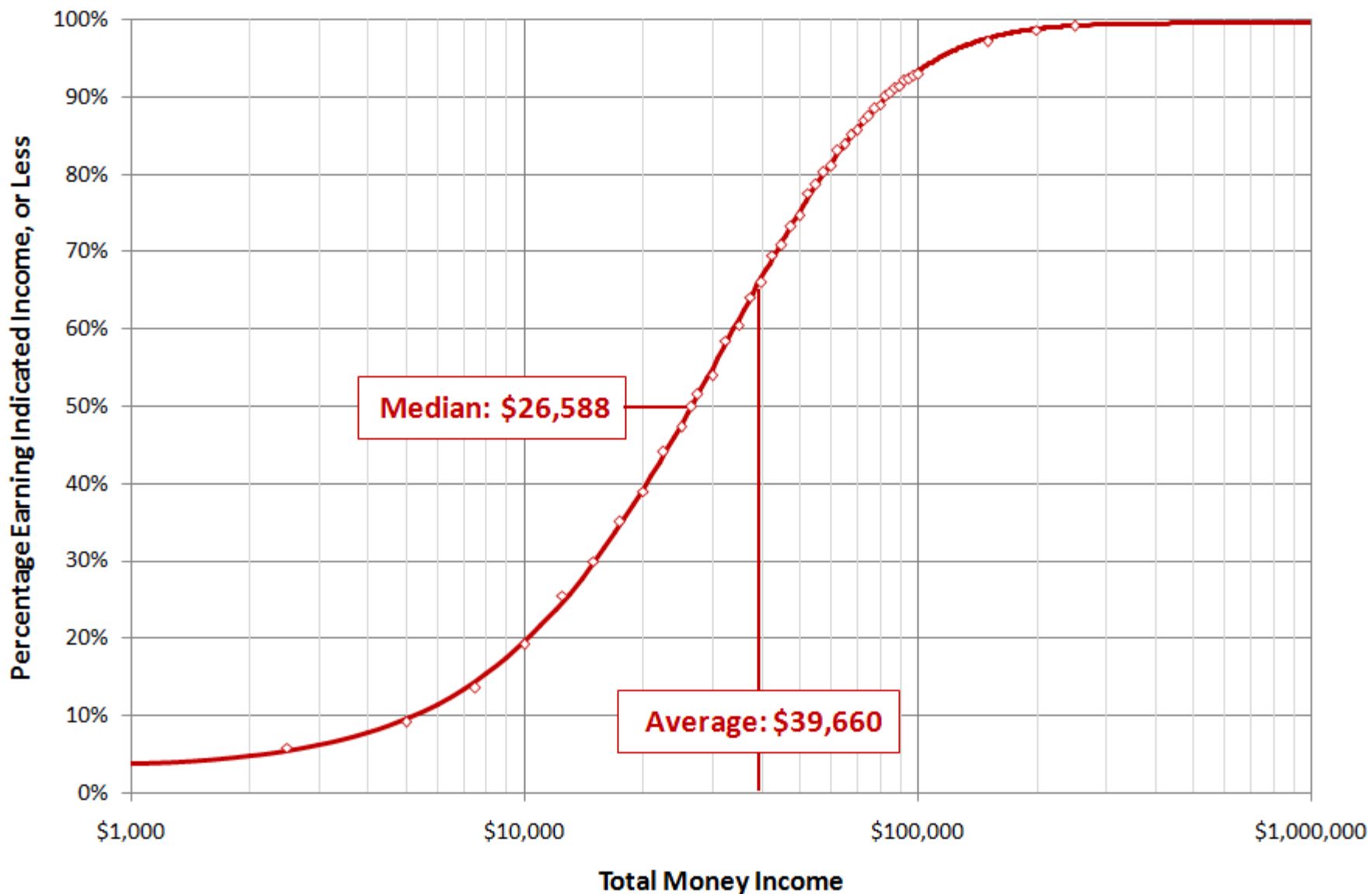
SELF-REPORTED HIGH SCHOOL GRADES	ENGLISH	MATH	FOREIGN LANGUAGE	BIOLOGICAL SCIENCES	PHYSICAL SCIENCES	SOCIAL STUDIES	AVERAGE OF THESE GRADES
	LATEST GRADES	B	B	A	B	C	B
HONORS COURSES							TOTAL EXPECTED YEARS OF STUDY
EXPECTED YEARS OF STUDY	5	3	3	1	2	3	17

SCORES		SAT					ACHIEVEMENT TESTS			
CURRENT SCORES & PERCENTILES		VERBAL	VERBAL SUBSCORES		MATH	TSWE	ACH 1	ACH 2	ACH 3	ACH AVG
TEST DATE	GRADE LEVEL	READING	WRITING	VOGAB. CLAR.						
NOV 84	12	700	68	71	640	58				
PERCENTILES										
NATIONAL H S SAMPLE		99			97					
COLLEGE-BOUND SENIORS		99	98	99	90	93				
PREVIOUS SCORES										
MAY 84	10	680	65	70	570	59				

PART-TIME
HRS WORKED
UNDER 6

ACHIEVEMENT
TEST CODES
AH: American History
& Social Studies

Cumulative Distribution of Total Money Income for U.S. Individuals, 2011



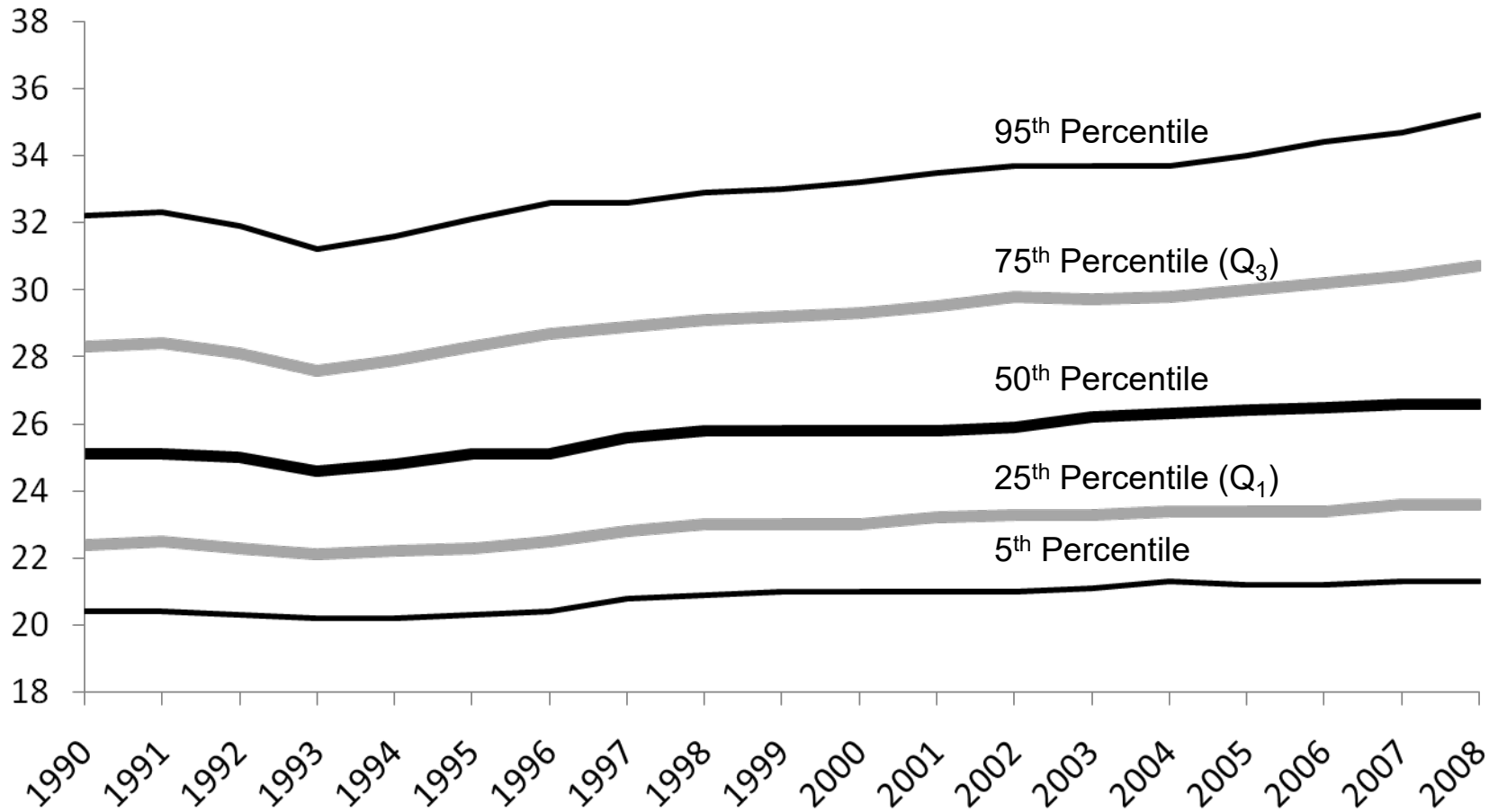
Percentiles

Percentiles are a more general form of the “five number summary” we talked about last time

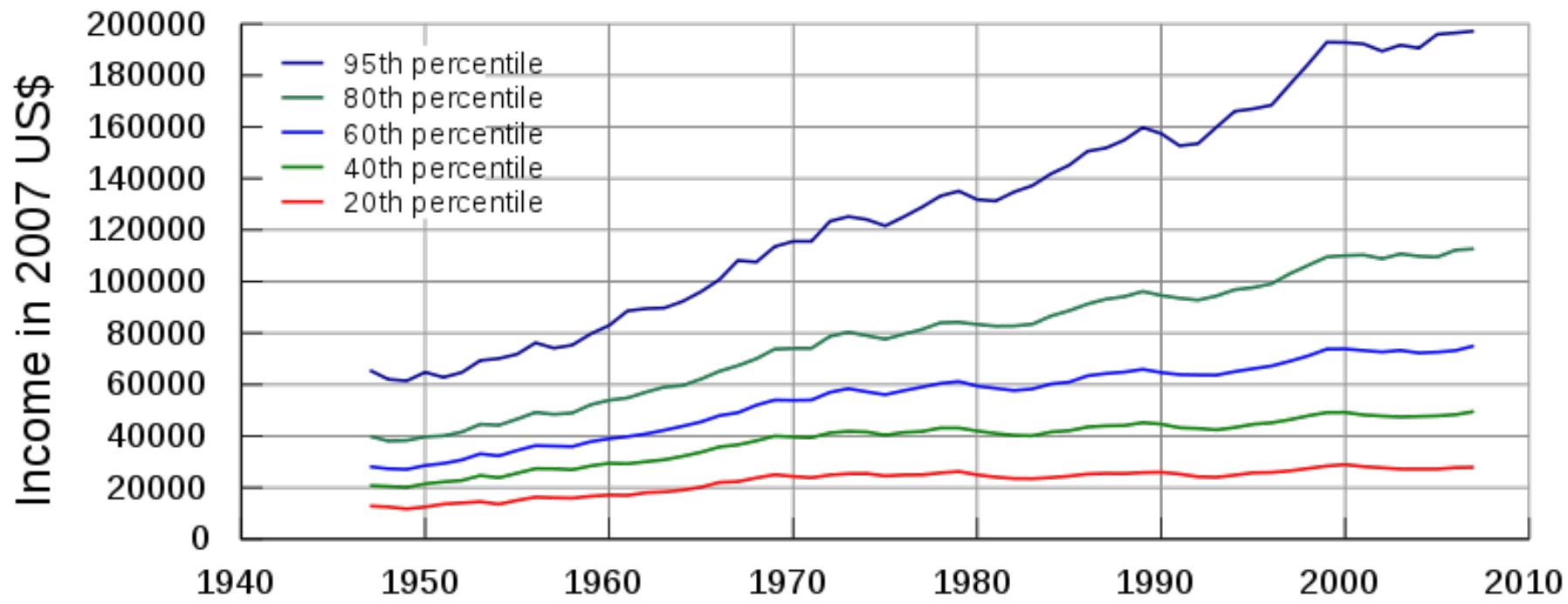
Minimum	→	0 th Percentile
Q_1	→	25 th Percentile
Median	→	50 th Percentile
Q_3	→	75 th Percentile
Maximum	→	100 th Percentile

Why *these* values? Other substantive applications may call for different “cut points”

Body Mass Index, 1990-2008



Source: 1990-2008 Integrated Health Interview Surveys



Percentiles

Quantiles

Divisions of the distribution into groups with known (and equal) proportions in each group

Examples:

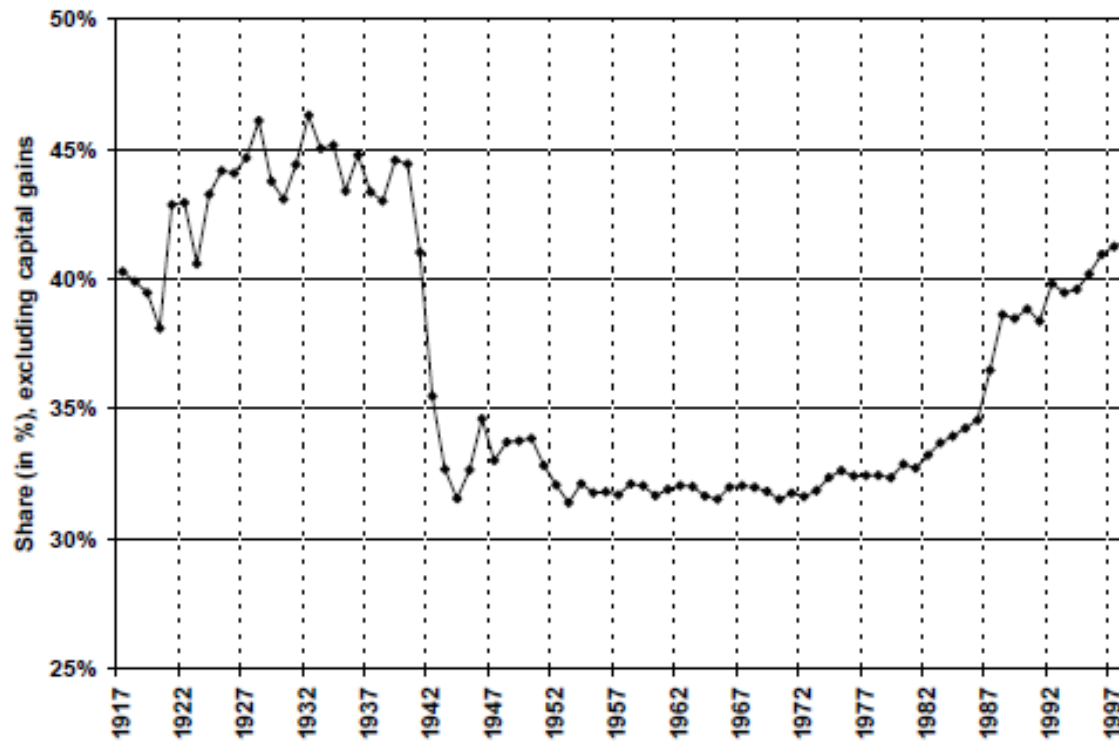
Quartiles = divisions of the distribution into four equal sized groups at the 25th percentile, the 50th percentile, and the 75th percentile

Quintiles = divisions of the distribution into five equal sized groups at the 20th percentile, the 40th percentile, etc.

Deciles = divisions of the distribution into ten equal sized groups at the 10th percentile, the 20th percentile, the 30th percentile, etc.

Percentiles

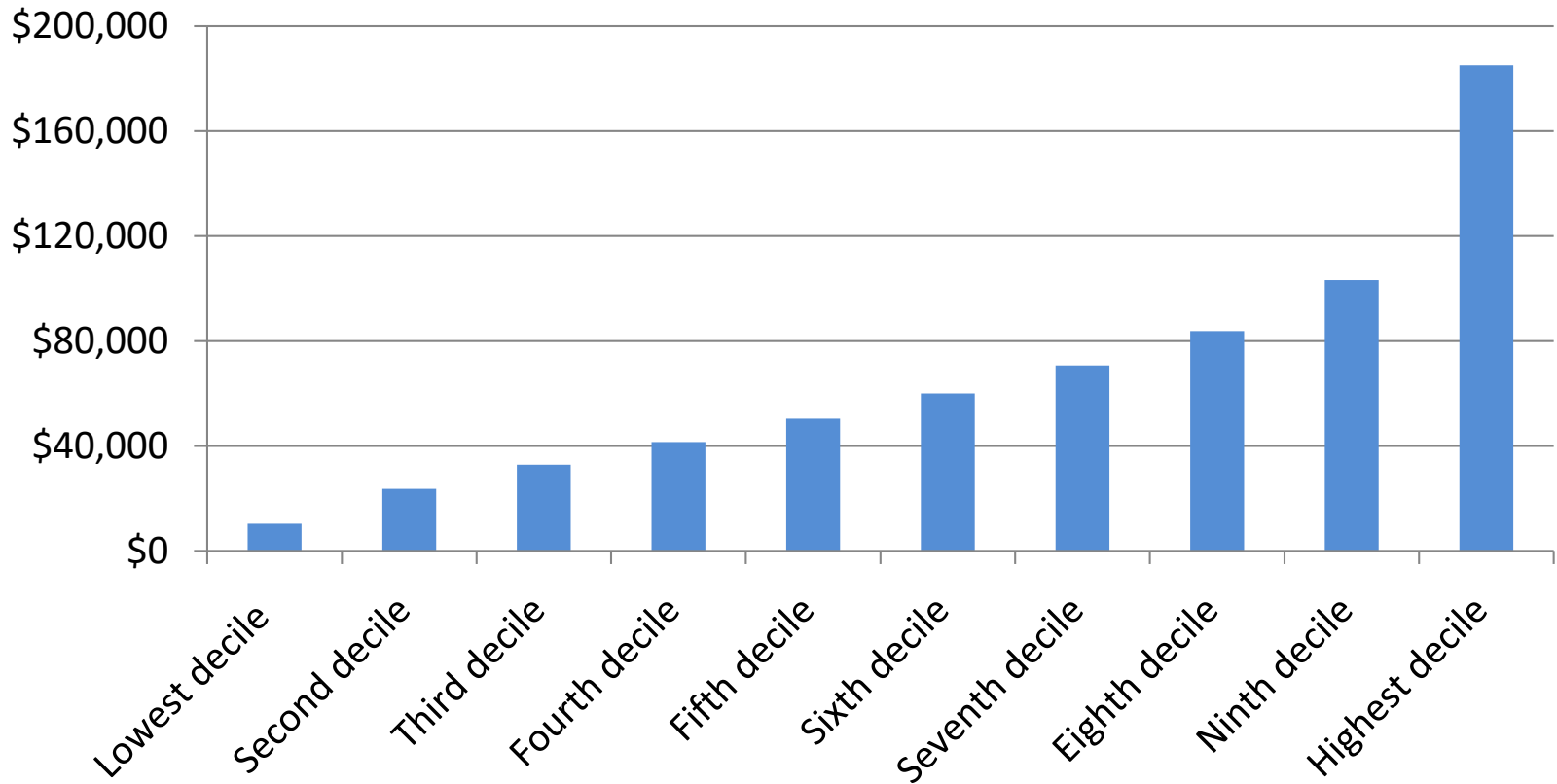
Figure 1: The top decile income share in the U.S., 1917-1998



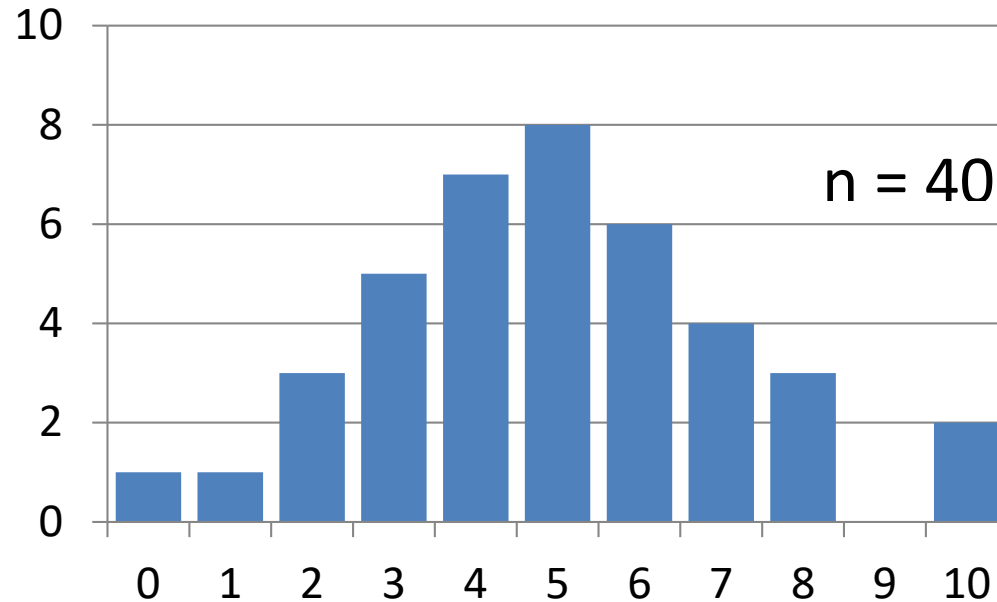
Source: Authors' computations based on Income tax returns (table A1, col. P90-100)

Percentiles

Average income, by income deciles, Canada, 2000



Worksheet



1. What is the 23rd percentile of this distribution?
2. What is the 99th percentile of this distribution?
3. If you wanted to divide this distribution into quintiles—five even sized groups—what values of the distribution would you use to separate the quartiles?

Change in Topic...



Standardized (Z) Scores

Standardized (Z) Score

“A transformation of the observed values of a continuous variable accomplished by subtracting the mean from each value and dividing by the standard deviation”

$$Z_i = \frac{(Y_i - \bar{Y})}{s_Y}$$

Example: Imagine a distribution of Y with a mean of 10 and a standard deviation of 4. How many standard deviations from the mean is a case with Y = 5?

$$Z_i = \frac{(5 - 10)}{4} = -1.25$$

Standardized (Z) Scores

$Z > 0$ → Observation has a value of Y above the mean

$Z = 0$ → Observation has a value of Y equal to the mean

$Z < 0$ → Observation has a value of Y below the mean

“Standardizing” a distribution by converting all of the observed values of Y to Z-scores results in a “standardized distribution” with a mean of 0 and a standard deviation of 1

Standardized (Z) Scores

Because all standardized distributions have a mean of 0 and a standard deviation of 1, it becomes possible to compare values across distributions that have different means and standard deviations

Example: Who is more unusually tall: A man who is 72 inches tall or a woman who is 67 inches tall?

$$\bar{Y}_{Men} = 69.1'' \text{ and } s_{Men} = 2.8$$

$$\bar{Y}_{Women} = 63.8'' \text{ and } s_{women} = 2.6$$

$$Z_{72''Man} = \frac{(72.0 - 69.1)}{2.8} = 1.04$$

$$Z_{67''Woman} = \frac{(67.0 - 63.8)}{2.6} = 1.23$$

Worksheet

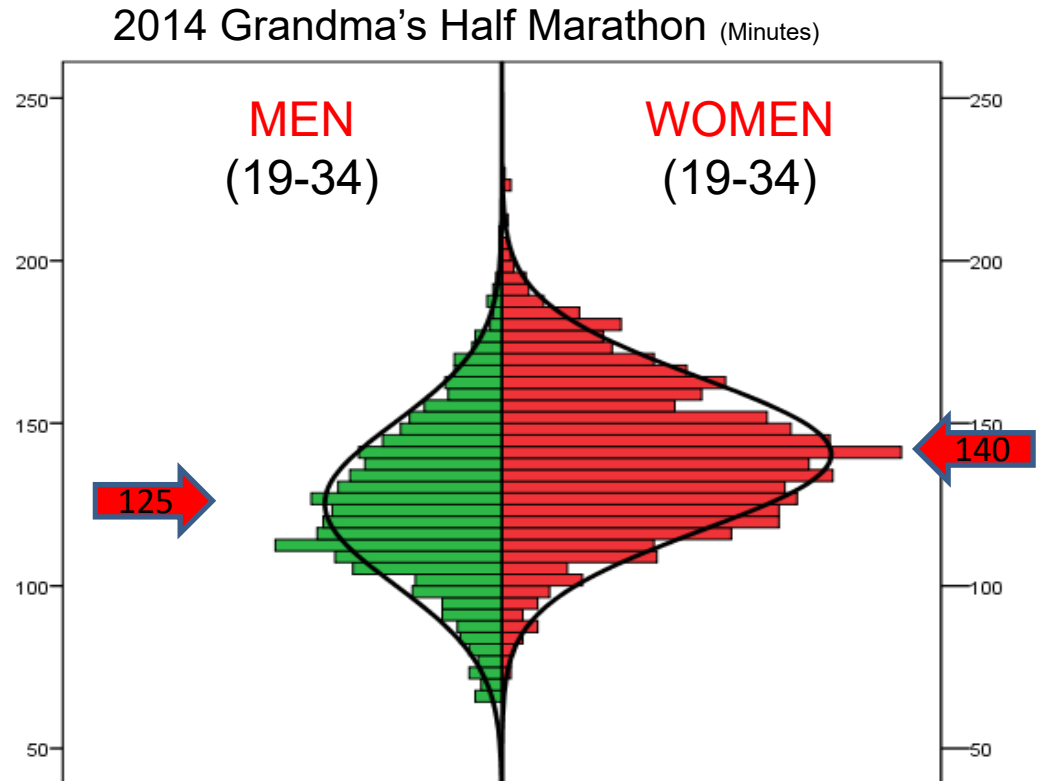
Who is more unusual? A man finishing the race in 109 minutes or a woman finishing in 124 minutes?

$$\bar{Y}_{Men} = 125$$

$$s_{Men} = 15$$

$$\bar{Y}_{Women} = 140$$

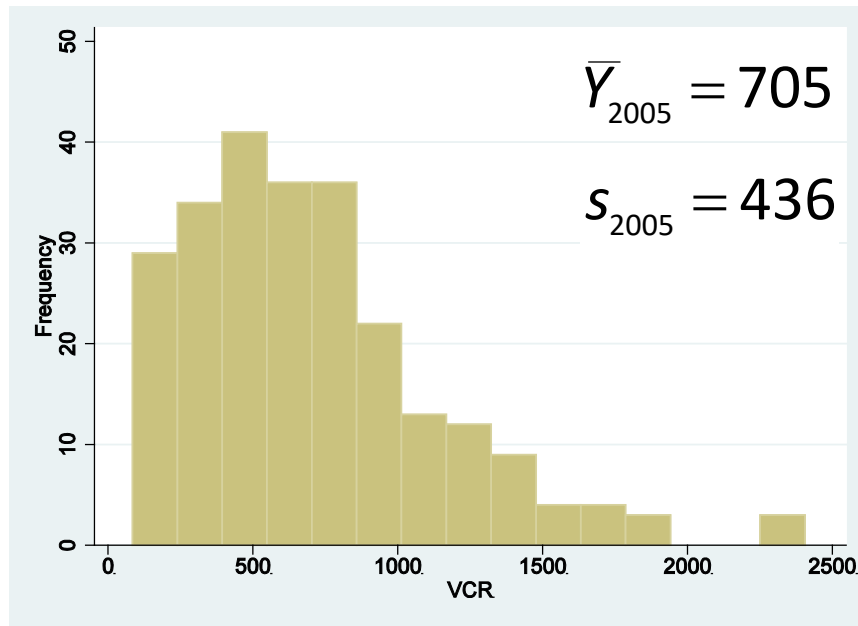
$$s_{women} = 11$$



Standardized (Z) Scores

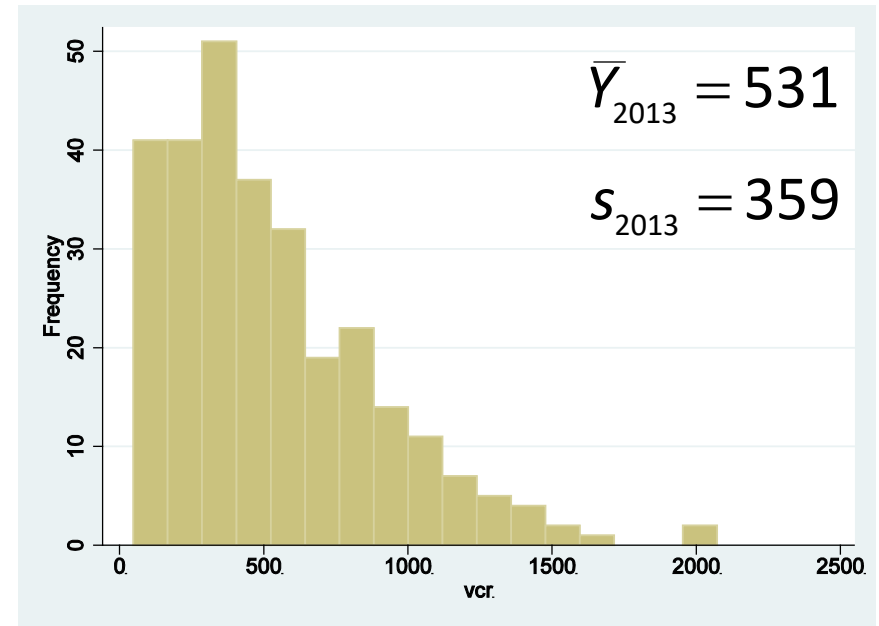
Here are violent crime rates (violent crimes per 100,000 residents) in 2005 and 2013 for cities in America with at least 100,000 people (from the Uniform Crime Reports)

2005



Minneapolis 1454
St Paul 876

2013



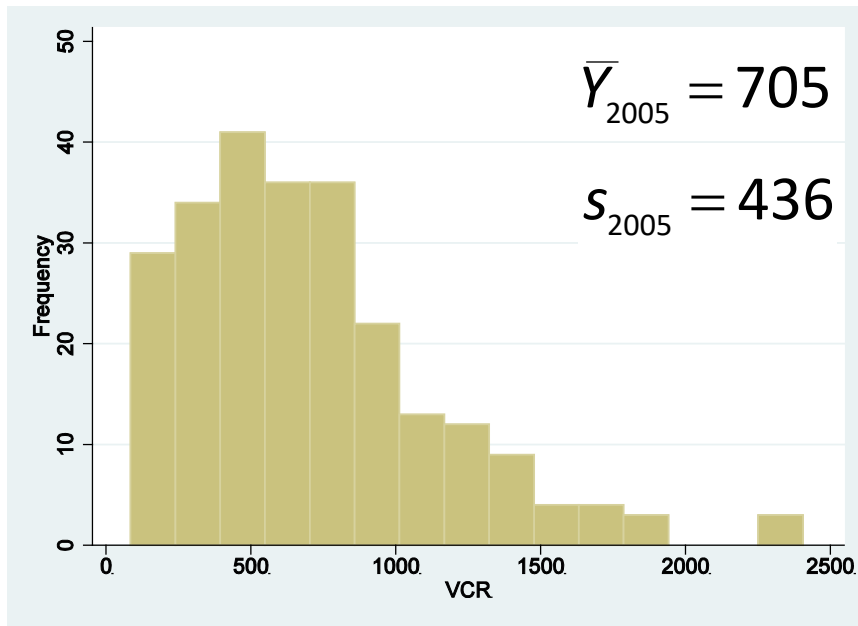
Minneapolis 1019
St Paul 747

How happy should Minneapolis and St Paul be about how much their violent crime rates decline between 2005 and 2013?

Standardized (Z) Scores

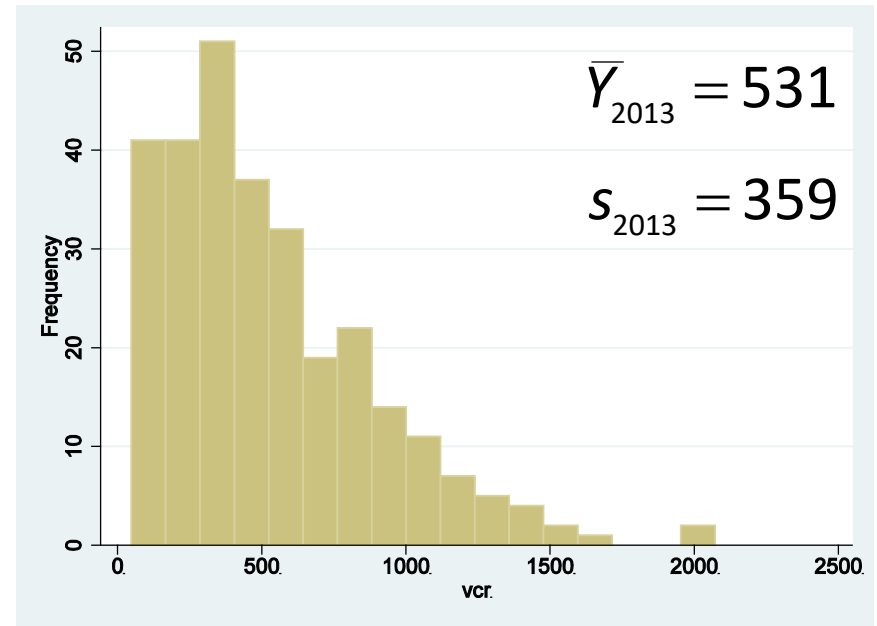
Here are violent crime rates (violent crimes per 100,000 residents) in 2005 and 2013 for cities in America with at least 100,000 people (from the Uniform Crime Reports)

2005



Minneapolis $Z=1.72$
St Paul $Z=0.39$

2013



Minneapolis $Z=1.36$
St Paul $Z=0.60$

Both saw declines in violent crime rates, but Minneapolis's rate declined faster than the national average while St Paul's declined slower than the national average

Worksheet

Mean Monthly Temperature in St Paul, MN, 1957-2006

What is more unusual? A January that averages 25 degrees or a September that averages 65 degrees?

