Detailed Overview of Variables and Potential Use

Autogenerated data summary from dataMaid

2019-01-05 21:58:49

Part 1

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	18883
Number of variables	44

- Only the following variables in fullData were included: kshare_2, capshare, familyFarm, urbanPop, nonagriPop, agriWorkforce, industWorkforce, labordepAgri, pcWealth, pcFinWealth, pcNonfinWealth, pcDebt, pcWealth.median, incomePre_p0p50, incomePre_p0p90, incomePre_p90p100, incomePre_p95p100, incomePre_p99p100, incomePost_p0p50, incomePost_p0p90, incomePost_p90p100, incomePost_p95p100, incomePost_p99p100, wealth_p0p50, wealth_p0p90, wealth_p90p100, wealth_p95p100, wealth_p99p100, wealth.top1, wealth.top10, wealth.top5, capshare.gross, capshare.net, lwheatsugar, stir, ltrate, stocks, debtgdp, tloans, tmort, thh, tbus, hpnom

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	haven labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×	×				
Identify levels with < 6 obs.	×	×	×	×				
Identify case issues	×	×	×	×				
Identify misclassified numeric or integer variables	×	×	×	×				
Identify outliers					×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

Part 2

Summary table

	Variable class	# unique values	Missing observations	Any problems?
kshare_2	numeric	318	82.46 %	×
capshare	numeric	347	81.75 %	
capshare.gross	numeric	1732	90.55 %	×
capshare.net	numeric	1598	91.48 %	×
familyFarm	integer	66	92.85 %	×
urbanPop	integer	70	92.85 %	
nonagriPop	integer	63	92.85 %	
agriWorkforce	numeric	911	78.03 %	
industWorkforce	numeric	496	79.17 %	
labordepAgri	numeric	3330	82.35 %	
pcWealth	numeric	2537	85.65 %	×
pcFinWealth	numeric	2327	85.65 %	×
pcNonfinWealth	numeric	2476	85.65 %	×
pcDebt	numeric	1703	85.65 %	×
pcWealth.median	numeric	2405	85.65 %	×
incomePre_p0p50	numeric	400	97.71 %	×
incomePre_p0p90	numeric	427	97.56 %	
incomePre_p90p100	numeric	1073	94.09 %	×
incomePre_p95p100	numeric	1167	93.59 %	×
incomePre_p99p100	numeric	1317	92.74 %	×
incomePost_p0p50	numeric	324	98.11 %	×
incomePost_p0p90	numeric	302	98.22 %	
incomePost_p90p100	numeric	947	94.75 %	×
incomePost_p95p100	numeric	1041	94.26 %	×
incomePost_p99p100	numeric	1192	93.41 %	×
wealth_p0p50	numeric	213	98.69 %	×
wealth_p0p90	numeric	365	97.88 %	
wealth_p90p100	numeric	367	97.88 %	
wealth_p95p100	numeric	365	97.88 %	
wealth_p99p100	numeric	365	97.87 %	
wealth.top10	numeric	156	98.83 %	
wealth.top5	numeric	173	98.76 %	
wealth.top1	numeric	210	98.43 %	×
lwheatsugar	numeric	122	43.82 %	×
stir	numeric	1292	88.17 %	×

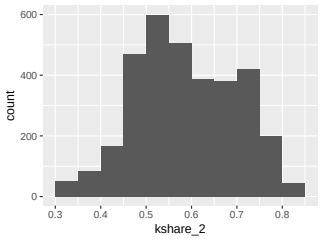
	Variable class	# unique values	Missing observations	Any problems?
ltrate	numeric	1689	87.58 %	×
stocks	numeric	2068	88.59 %	×
debtgdp	numeric	2179	88.10 %	×
tloans	numeric	2170	88.40 %	×
tmort	numeric	2073	88.81 %	×
thh	numeric	1221	93.51 %	×
tbus	numeric	1147	93.91 %	×
hpnom	numeric	1738	90.28 %	

Part 3

Variable list

kshare_2

Feature	Result
Variable type	numeric
Number of missing obs.	15570 (82.46 %)
Number of unique values	317
Median	0.57
1st and 3rd quartiles	0.51; 0.68
Min. and max.	0.32; 0.85

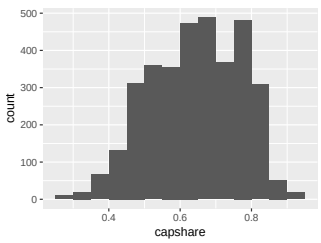


- Note that the following possible outlier values were detected: "0.32", "0.32", "0.34", "0.34", "0.34", "0.36", "0.36".
- Variable Description: variable measures capital share minus oil rents
- Source: Dunning, 1960–2001
- How to use: theoretical relationship between capital share and wealth inequality not entirely clear; capital share can be result of (a) productivity of capital, (b) demand elasticity between capital and labor, and (c) strength (or rather weakness) of labor market institutions; it is thus a very rough measure for power of capital
- Issues: only post 1960

---

capshare

Feature	Result
Variable type	numeric
Number of missing obs.	15436 (81.75 %)
Number of unique values	346
Median	0.65
1st and 3rd quartiles	0.54; 0.75
Min. and max.	0.29; 0.93

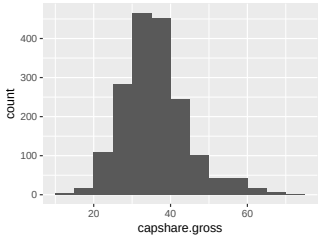


- Variable Description: variable measures capital share
- Source: Houle, 1960–2000
- How to use: theoretical relationship between capital share and wealth inequality not entirely clear; capital share can be result of (a) productivity of capital, (b) demand elasticity between capital and labor, and (c) strength (or rather weakness) of labor market institutions; it is thus a very rough measure for power of capital
- Issues: only post 1960

---

capshare.gross

Feature	Result
Variable type	numeric
Number of missing obs.	17099 (90.55 %)
Number of unique values	1731
Median	35.09
1st and 3rd quartiles	30.55; 40.15
Min. and max.	11.73; 70.9

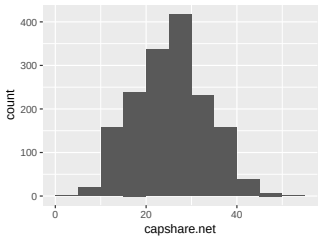


- Note that the following possible outlier values were detected: "11.73", "12.46", "14.88", "14.96", "16.71", "17.18", "17.62", "17.85", "17.99", "18.38" (43 additional values omitted).
- Variable Description: variable measures gross capital share
- Source: Bengtsson & Waldenstroem, 1875–2015
- How to use: theoretical relationship between capital share and wealth inequality not entirely clear; capital share can be result of (a) productivity of capital, (b) demand elasticity between capital and labor, and (c) strength (or rather weakness) of labor market institutions; it is thus a very rough measure for power of capital; might be superior to `capshare` and `k_share2` because of longer time period covered
- Issues: need to check correlation with `capshare` and `k_share2`

---

capshare.net

Feature	Result
Variable type	numeric
Number of missing obs.	17274 (91.48 %)
Number of unique values	1597
Median	25.53
1st and 3rd quartiles	19.67; 30.65
Min. and max.	1.38; 52.24



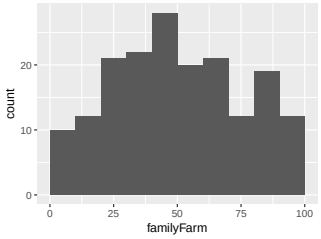
- Note that the following possible outlier values were detected: "1.38", "1.44", "45.55", "45.64", "45.78", "45.96", "46.94", "47.18", "47.47", "50.96" (1 additional values omitted).
- Variable Description: variable measures gross capital share net of capital depreciation
- Source: Bengtsson & Waldenstroem, 1875–2015

- How to use: theoretical relationship between capital share and wealth inequality not entirely clear; capital share can be result of (a) productivity of capital, (b) demand elasticity between capital and labor, and (c) strength (or rather weakness) of labor market institutions; it is thus a very rough measure for power of capital; might be superior to `capshare` and `k_share2` because of longer time period covered
- Issues: need to check correlation with `capshare` and `k_share2`

---

familyFarm

Feature	Result
Variable type	integer
Number of missing obs.	2300 (92.85 %)
Number of unique values	65
Median	48
1st and 3rd quartiles	33; 70
Min. and max.	1; 98

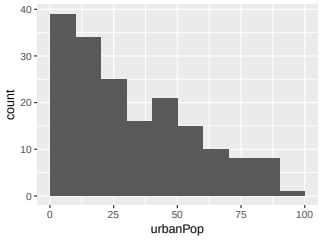


- Variable Description: variable measures proportion of all cultivable land worked by families
- Source: Vanhanen, 1858-2000
- How to use: higher values indicate lower agricultural concentration, i.e. rough proxy for inequality especially in 19th century.
- Issues: measured only every 10 years (needs to be interpolated).

---

urbanPop

Feature	Result
Variable type	integer
Number of missing obs.	2300 (92.85 %)
Number of unique values	69
Median	27
1st and 3rd quartiles	13; 49
Min. and max.	3; 91

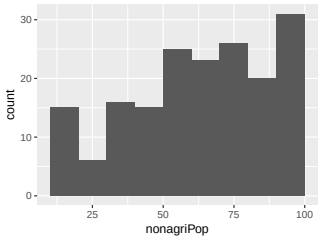


- Variable Description: variable measures percentage of urban population
- Source: Vanhanen, 1858-2000
- How to use: can help to measure to what extent inequality measures based on agricultural inequality are adequate proxies
- Issues: measured only every 10 years (needs to be interpolated)

---

nonagriPop

Feature	Result
Variable type	integer
Number of missing obs.	2300 (92.85 %)
Number of unique values	62
Median	65
1st and 3rd quartiles	46; 84
Min. and max.	10; 98



- Variable Description: variable measures percentage of population in non-agricultural occupation
- Source: Vanhanen, 1858-2000
- How to use: measure to what extent inequality measures based on agricultural inequality are adequate proxies. Need to check how this is correlated with measures of urbanization (`urbanPop`).
- Issues: measured only every 10 years (needs to be interpolated)

---

agriWorkforce

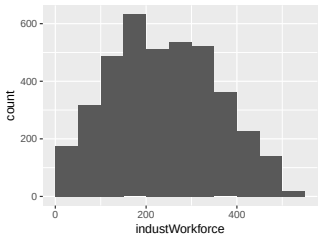
Feature	Result
Variable type	numeric
Number of missing obs.	14735 (78.03 %)
Number of unique values	910
Median	451
1st and 3rd quartiles	264.75; 605
Min. and max.	11; 969

- Variable Description: variable measures percentage of workforce working in agriculture
- Source: Banks & Wilson (CNTS), 1815–1980
- How to use: measure to what extent inequality measures based on agricultural inequality are adequate proxies. Need to check how this is correlated with measures for industrial workforce (`industWorkforce`)
- Issues: Cannot be used for most recent years (post-1980). Need to investigate why not bounded by 0 and 100.

---

industWorkforce

Feature	Result
Variable type	numeric
Number of missing obs.	14950 (79.17 %)
Number of unique values	495
Median	237
1st and 3rd quartiles	151; 328
Min. and max.	0; 523

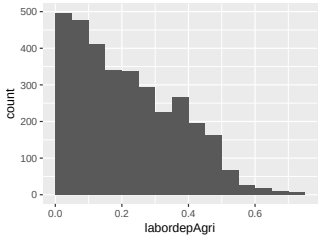


- Variable Description: variable measures percentage of workforce working in industry
- Source: Banks & Wilson (CNTS), 1840–1980
- How to use: measure to what extent inequality measures based on agricultural inequality are adequate proxies because it measures degree of industrialization. Need to check how this is correlated with measures for agricultural workforce (`agriWorkforce`).
- Issues: Cannot be used for most recent years (post-1980). Need to investigate why not bounded by 0 and 100.

---

labordepAgri

Feature	Result
Variable type	numeric
Number of missing obs.	15550 (82.35 %)
Number of unique values	3329
Median	0.19
1st and 3rd quartiles	0.08; 0.33
Min. and max.	0; 0.73

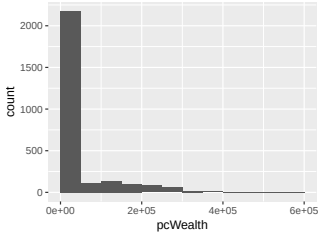


- Variable Description: variable measures share of population working in agriculture that is not landowner itself and therefore relies on employers/landlords for access to land
- Source: Albertus, 1930–2006
- How to use: measure of inequality, but only in contexts where wealth was mainly based on landholdings
- Issues: data is only available post-1930, i.e. when most Western countries had industrialized already and land inequality is less of a good measure to proxy wealth inequality

---

pcWealth

Feature	Result
Variable type	numeric
Number of missing obs.	16173 (85.65 %)
Number of unique values	2536
Median	9856
1st and 3rd quartiles	2829.5; 30271.25
Min. and max.	116; 552764



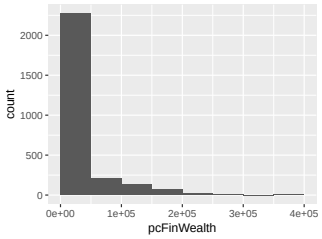
- Note that the following possible outlier values were detected: "275063", "275306", "275644", "276308", "277271", "278435", "279236", "279409", "280922", "281572" (65 additional values omitted).
- Variable Description: variable measures mean absolute wealth per adult
- Source: Credit Suisse, 2000–2017
- How to use: extend analyses to more recent years, but needs to be combined with variable for median wealth per adult (`pcWealth.median`); perhaps need get to net wealth by subtract mean absolute debt per adult (`pcDebt`)

- Issues: only available post-2000; data from Credit Suisse database is collected in weird way by simply extrapolating data for countries from neighboring countries

---

pcFinWealth

Feature	Result
Variable type	numeric
Number of missing obs.	16173 (85.65 %)
Number of unique values	2326
Median	2898
1st and 3rd quartiles	914.5; 14927
Min. and max.	5; 383908

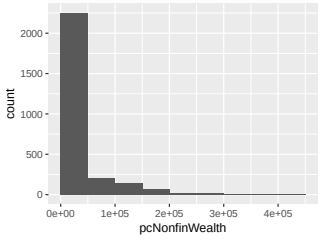


- Note that the following possible outlier values were detected: "229351", "230842", "232813", "232962", "234618", "236328", "242800", "248492", "252905", "263304" (12 additional values omitted).
- Variable Description: variable measures mean absolute financial wealth per adult
- Source: Credit Suisse, 2000–2017
- How to use: measure of ratio of financial wealth to total wealth, combine with measure of mean total wealth (`pcWealth`)
- Issues: only available post-2000; data from Credit Suisse database is collected in weird way by simply extrapolating data for countries from neighboring countries; check if `pcFinWealth` and `pcNonFinWealth` sum up to `pcWealth`

---

pcNonfinWealth

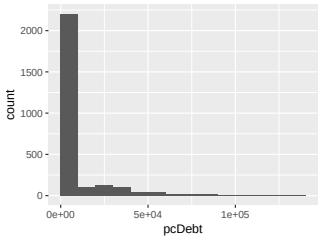
Feature	Result
Variable type	numeric
Number of missing obs.	16173 (85.65 %)
Number of unique values	2475
Median	7422
1st and 3rd quartiles	1707.5; 21236
Min. and max.	46; 416495



- Note that the following possible outlier values were detected: "166071", "166201", "167023", "167211", "167295", "169048", "169542", "169903", "170677", "172009" (70 additional values omitted).
- Variable Description: variable measures mean absolute non-financial wealth per adult
- Source: Credit Suisse, 2000–2017
- How to use: measure of ratio of non-financial wealth to total wealth, combine with measure of mean total wealth (`pcWealth`)
- Issues: only available post-2000; data from Credit Suisse database is collected in weird way by simply extrapolating data for countries from neighboring countries

pcDebt

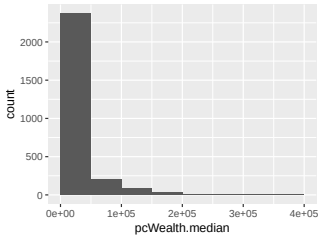
Feature	Result
Variable type	numeric
Number of missing obs.	16173 (85.65 %)
Number of unique values	1702
Median	568
1st and 3rd quartiles	95; 4983.25
Min. and max.	1; 132825



- Note that the following possible outlier values were detected: "97425", "98041", "98446", "100783", "102177", "102205", "102770", "103023", "103713", "106446" (17 additional values omitted).
- Variable Description: variable measures mean absolute debt per adult
- Source: Credit Suisse, 2000–2017
- How to use: calculate net wealth by subtracting from mean total wealth per adult (`pcWealth`)
- Issues: only available post-2000; data from Credit Suisse database is collected in weird way by simply extrapolating data for countries from neighboring countries

pcWealth.median

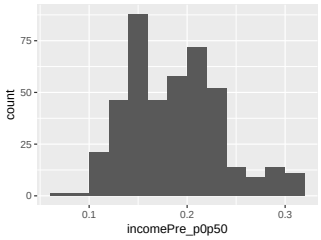
Feature	Result
Variable type	numeric
Number of missing obs.	16173 (85.65 %)
Number of unique values	2404
Median	4051
1st and 3rd quartiles	1231.75; 14205.25
Min. and max.	46; 397767



- Note that the following possible outlier values were detected: "153905", "154044", "154579", "154805", "156294", "157158", "157480", "157530", "160382", "161334" (42 additional values omitted).
- Variable Description: variable measures median absolute debt per adult
- Source: Credit Suisse, 2000–2017
- How to use: measure is less sensitive to outliers; by combining it with mean total wealth per adult (`pcWealth`) we can obtain measure of skewness of wealth distribution
- Issues: only available post-2000; data from Credit Suisse database is collected in weird way by simply extrapolating data for countries from neighboring countries

incomePre_p0p50

Feature	Result
Variable type	numeric
Number of missing obs.	18450 (97.71 %)
Number of unique values	399
Median	0.19
1st and 3rd quartiles	0.15; 0.22
Min. and max.	0.08; 0.32

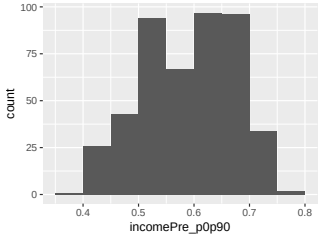


- Note that the following possible outlier values were detected: "0.3", "0.3", "0.3", "0.3", "0.3", "0.31", "0.31", "0.31", "0.31", "0.32" (1 additional values omitted).
- Variable Description: variable measures pre-fiscal income share of bottom 50%
- Source: World Income Database, 1900–2015
- How to use: measure of income inequality; proxy for share of income claimed by poor
- Issues:

---

incomePre_p0p90

Feature	Result
Variable type	numeric
Number of missing obs.	18423 (97.56 %)
Number of unique values	426
Median	0.6
1st and 3rd quartiles	0.53; 0.66
Min. and max.	0.39; 0.76

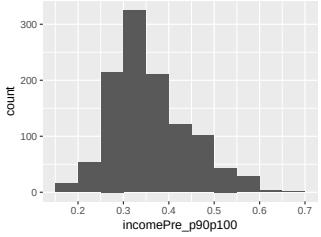


- Variable Description: variable measures pre-fiscal income share of bottom 90%
- Source: World Income Database, 1900–2015
- How to use: measure of income inequality; proxy for share of income claimed by poor and middle class
- Issues:

---

incomePre_p90p100

Feature	Result
Variable type	numeric
Number of missing obs.	17767 (94.09 %)
Number of unique values	1072
Median	0.34
1st and 3rd quartiles	0.3; 0.41
Min. and max.	0.15; 0.65

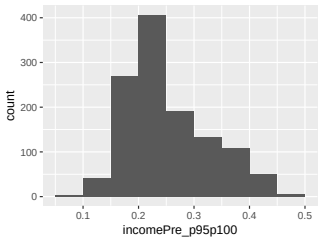


- Note that the following possible outlier values were detected: "0.15", "0.16", "0.17", "0.17", "0.18", "0.18", "0.18", "0.18", "0.18", "0.19" (22 additional values omitted).
- Variable Description: variable measures pre-fiscal income share of top 10%
- Source: World Income Database, 1875–2015
- How to use: measure of income inequality; proxy for share of income claimed by rich
- Issues:

---

incomePre_p95p100

Feature	Result
Variable type	numeric
Number of missing obs.	17673 (93.59 %)
Number of unique values	1166
Median	0.23
1st and 3rd quartiles	0.2; 0.3
Min. and max.	0.09; 0.49

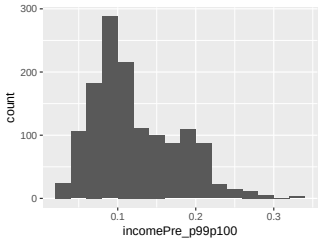


- Note that the following possible outlier values were detected: "0.09", "0.09", "0.1", "0.1", "0.1", "0.1", "0.1", "0.11", "0.11", "0.11" (44 additional values omitted).
- Variable Description: variable measures pre-fiscal income share of top 5%
- Source: World Income Database, 1888-2015
- How to use: measure of income inequality; proxy for share of income claimed by rich
- Issues:

---

incomePre_p99p100

Feature	Result
Variable type	numeric
Number of missing obs.	17513 (92.74 %)
Number of unique values	1316
Median	0.11
1st and 3rd quartiles	0.08; 0.16
Min. and max.	0.02; 0.33



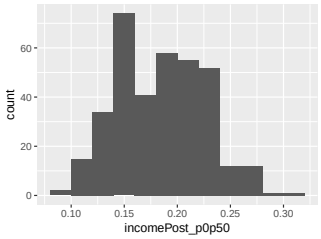
- Note that the following possible outlier values were detected: "0.02", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03" (66 additional values omitted).
- Variable Description: variable measures pre-fiscal income share of top 1%
- Source: World Income Database, 1875–2015
- How to use: measure of income inequality; proxy for share of income claimed by very rich

- Issues:

---

incomePost_p0p50

Feature	Result
Variable type	numeric
Number of missing obs.	18526 (98.11 %)
Number of unique values	323
Median	0.18
1st and 3rd quartiles	0.15; 0.22
Min. and max.	0.1; 0.32

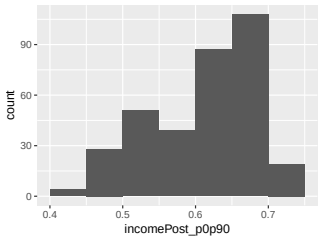


- Note that the following possible outlier values were detected: "0.32".
- Variable Description: variable measures post-fiscal income share of bottom 50%
- Source: World Income Database, 1900–2015
- How to use: measure of income inequality; proxy for share of income claimed by poor
- Issues:

---

incomePost_p0p90

Feature	Result
Variable type	numeric
Number of missing obs.	18547 (98.22 %)
Number of unique values	301
Median	0.63
1st and 3rd quartiles	0.55; 0.68
Min. and max.	0.45; 0.74

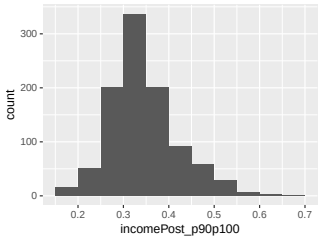


- Variable Description: variable measures post-fiscal income share of bottom 90%
- Source: World Income Database, 1900–2015
- How to use: measure of income inequality; proxy for share of income claimed by poor and middle class
- Issues:

---

incomePost_p90p100

Feature	Result
Variable type	numeric
Number of missing obs.	17892 (94.75 %)
Number of unique values	946
Median	0.33
1st and 3rd quartiles	0.3; 0.38
Min. and max.	0.15; 0.65

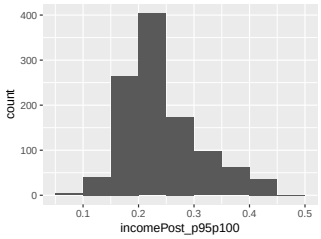


- Note that the following possible outlier values were detected: "0.15", "0.16", "0.17", "0.17", "0.18", "0.18", "0.18", "0.18", "0.18", "0.19" (31 additional values omitted).
- Variable Description: variable measures post-fiscal income share of top 10%
- Source: World Income Database, 1875–2015
- How to use: measure of income inequality; proxy for share of income claimed by rich; difference between `incomePre_p90p100` and `incomePost_p90p100` as rough measure of redistribution (esp. in 19th century when few measures of redistribution exist)
- Issues:

---

incomePost_p95p100

Feature	Result
Variable type	numeric
Number of missing obs.	17799 (94.26 %)
Number of unique values	1040
Median	0.23
1st and 3rd quartiles	0.2; 0.28
Min. and max.	0.09; 0.47

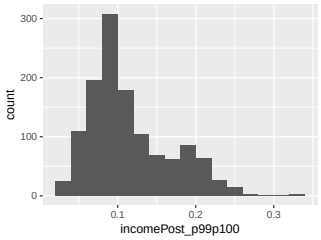


- Note that the following possible outlier values were detected: "0.09", "0.09", "0.1", "0.1", "0.1", "0.1", "0.1", "0.11", "0.11", "0.11" (48 additional values omitted).
- Variable Description: variable measures post-fiscal income share of top 5%
- Source: World Income Database, 1888–2015
- How to use: measure of income inequality; proxy for share of income claimed by rich; difference between `incomePre_p95p100` and `incomePost_p95p100` as rough measure of redistribution (esp. in 19th century when few measures of redistribution exist)
- Issues:

---

incomePost_p99p100

Feature	Result
Variable type	numeric
Number of missing obs.	17638 (93.41 %)
Number of unique values	1191
Median	0.1
1st and 3rd quartiles	0.08; 0.14
Min. and max.	0.02; 0.33

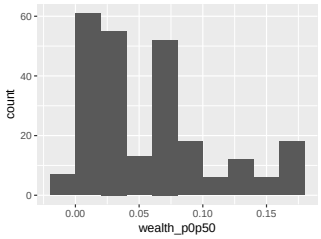


- Note that the following possible outlier values were detected: "0.02", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03", "0.03" (93 additional values omitted).
- Variable Description: variable measures post-fiscal income share of top 1%
- Source: World Income Database, 1875–2015
- How to use: measure of income inequality; proxy for share of income claimed by very rich; difference between `incomePre_p99p100` and `incomePost_p99p100` as rough measure of redistribution (esp. in 19th century when few measures of redistribution exist)
- Issues:

---

wealth_p0p50

Feature	Result
Variable type	numeric
Number of missing obs.	18635 (98.69 %)
Number of unique values	212
Median	0.04
1st and 3rd quartiles	0.02; 0.08
Min. and max.	-0.02; 0.16

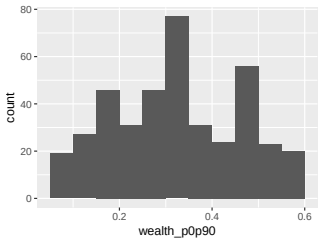


- Note that the following possible outlier values were detected: "-0.02", "-0.02", "-0.01", "-0.01", "-0.01", "0".
- Variable Description: variable measures net personal wealth share of bottom-50%
- Source: World Income Database, 1807–2015
- How to use: measure of wealth; proxy for share of wealth claimed by poor
- Issues: only data for 4 countries

---

wealth_p0p90

Feature	Result
Variable type	numeric
Number of missing obs.	18483 (97.88 %)
Number of unique values	364
Median	0.32
1st and 3rd quartiles	0.21; 0.45
Min. and max.	0.07; 0.59

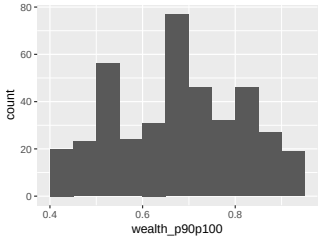


- Variable Description: variable measures net personal wealth share of bottom-90%
- Source: World Income Database, 1807–2015
- How to use: measure of wealth; proxy for share of wealth claimed by poor and middle class
- Issues: only data for 5 countries

---

wealth_p90p100

Feature	Result
Variable type	numeric
Number of missing obs.	18482 (97.88 %)
Number of unique values	366
Median	0.68
1st and 3rd quartiles	0.55; 0.79
Min. and max.	0.41; 0.93

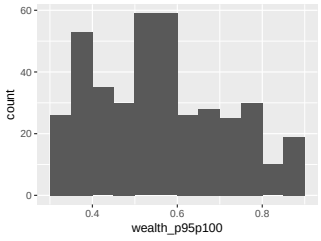


- Variable Description: variable measures net personal wealth share of top-10%
- Source: World Income Database, 1807–2015
- How to use: measure of wealth; proxy for share of wealth claimed by rich
- Issues: only data for 5 countries

---

wealth_p95p100

Feature	Result
Variable type	numeric
Number of missing obs.	18483 (97.88 %)
Number of unique values	364
Median	0.55
1st and 3rd quartiles	0.43; 0.67
Min. and max.	0.3; 0.9



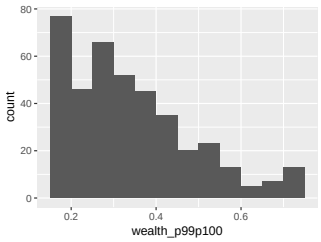
- Variable Description: variable measures net personal wealth share of top-5%
- Source: World Income Database, 1807–2015
- How to use: measure of wealth; proxy for share of wealth claimed by rich

- Issues: only data for 5 countries

---

wealth_p99p100

Feature	Result
Variable type	numeric
Number of missing obs.	18481 (97.87 %)
Number of unique values	364
Median	0.31
1st and 3rd quartiles	0.23; 0.43
Min. and max.	0.15; 0.74



- Variable Description: variable measures net personal wealth share of top-1%
- Source: World Income Database, 1807–2015
- How to use: measure of wealth; proxy for share of wealth claimed by very rich
- Issues: only data for 5 countries

---

wealth.top10

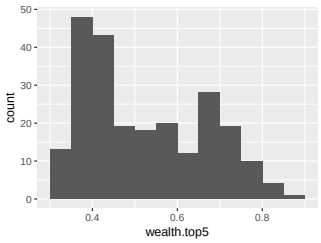
Feature	Result
Variable type	numeric
Number of missing obs.	18662 (98.83 %)
Number of unique values	155
Median	0.63
1st and 3rd quartiles	0.53; 0.8
Min. and max.	0.46; 0.92

- Variable Description: variable measures net personal wealth share of top-10%
- Source: Roine & Waldenstroem, 1800–2011
- How to use: measure of wealth; proxy for share of wealth claimed by rich
- Issues: data for 7 countries (adds 5 countries to WID data: CH; FI; SE; NO; DK)

---

wealth.top5

Feature	Result
Variable type	numeric
Number of missing obs.	18648 (98.76 %)
Number of unique values	172
Median	0.48
1st and 3rd quartiles	0.4; 0.66
Min. and max.	0.32; 0.87

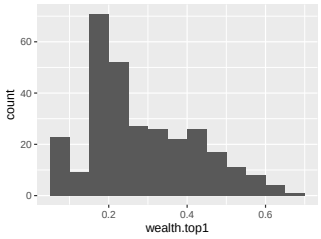


- Variable Description: variable measures net personal wealth share of top-5%
- Source: Roine & Waldenstroem, 1800–2011
- How to use: measure of wealth; proxy for share of wealth claimed by rich
- Issues: data for 7 countries (adds 5 countries to WID data: CH; FI; SE; NO; DK)

---

wealth.top1

Feature	Result
Variable type	numeric
Number of missing obs.	18586 (98.43 %)
Number of unique values	209
Median	0.24
1st and 3rd quartiles	0.19; 0.39
Min. and max.	0.06; 0.69

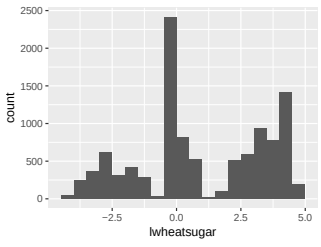


- Note that the following possible outlier values were detected: "0.06", "0.07", "0.07", "0.07", "0.07", "0.07", "0.08", "0.08", "0.08", "0.08" (18 additional values omitted).
- Variable Description: variable measures net personal wealth share of top-1%
- Source: Roine & Waldenstroem, 1800–2011
- How to use: measure of wealth; proxy for share of wealth claimed by very rich
- Issues: data for 9 countries (adds 7 countries to WID data: NL; CH; FI; SE; NO; DK; AU)

---

lwheatsugar

Feature	Result
Variable type	numeric
Number of missing obs.	8275 (43.82 %)
Number of unique values	121
Median	0.31
1st and 3rd quartiles	-0.2; 3.27
Min. and max.	-4.03; 4.56

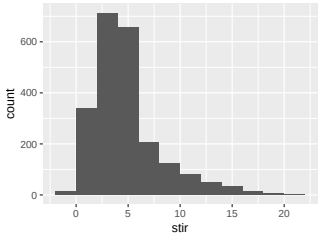


- Note that the following possible outlier values were detected: "-4.03", "-3.94", "-3.74", "-3.69", "-3.59", "-3.41", "-3.4", "-3.36", "-3.15", "-3.01" (23 additional values omitted).
- Variable Description: variable measures log of percent of land suitable for wheat to percent of land suitable for sugar crane
- Source: Easterly (used by Albertus), 1900–2009
- How to use: measure for elite bias rooted in agriculture as wheat cultivation on family farms is tied to egalitarian institutions while sugar crane cultivation on large plantations generated inegalitarian institutions (i.e. higher values indicate lower elite bias)
- Issues:

---

stir

Feature	Result
Variable type	numeric
Number of missing obs.	16649 (88.17 %)
Number of unique values	1291
Median	4.17
1st and 3rd quartiles	2.85; 5.88
Min. and max.	-2; 21.27

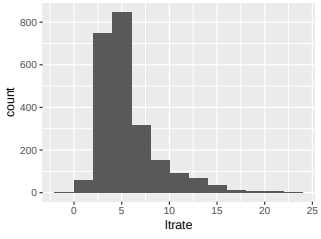


- Note that the following possible outlier values were detected: "-2", "-1", "-0.29", "-0.22", "-0.2", "-0.13", "-0.11", "-0.08", "-0.02", "-0.02" (148 additional values omitted).
- Variable Description: variable measures short-term interest rate (nominal; percent per year)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use:
- Issues:

---

ltrate

Feature	Result
Variable type	numeric
Number of missing obs.	16538 (87.58 %)
Number of unique values	1688
Median	4.64
1st and 3rd quartiles	3.64; 6.4
Min. and max.	-0.04; 23.72



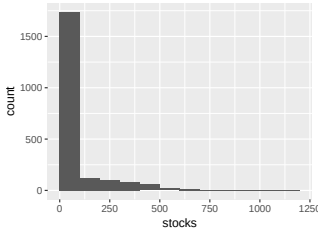
- Note that the following possible outlier values were detected: "-0.04", "0.35", "0.38", "0.5", "0.53", "0.56", "0.69", "0.69", "0.7", "0.72" (99 additional values omitted).
- Variable Description: variable measures long-term interest rate (nominal; percent per year)

- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use:
- Issues:

---

stocks

Feature	Result
Variable type	numeric
Number of missing obs.	16729 (88.59 %)
Number of unique values	2067
Median	8.1
1st and 3rd quartiles	2.86; 48.91
Min. and max.	0; 1184.15

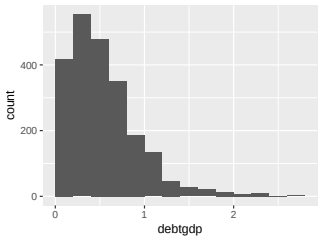


- Note that the following possible outlier values were detected: "0", "0", "0", "0", "0", "0", "0", "0", "0", "0" (128 additional values omitted).
- Variable Description: variable measures stock prices (nominal index)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use: measure to analyze whether abnormal stock returns are related with higher wealth inequality
- Issues:

---

debtgdp

Feature	Result
Variable type	numeric
Number of missing obs.	16636 (88.1 %)
Number of unique values	2178
Median	0.46
1st and 3rd quartiles	0.25; 0.7
Min. and max.	0.02; 2.7

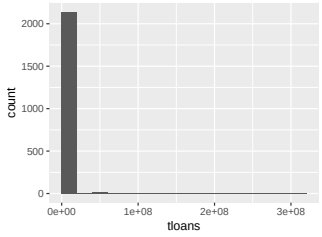


- Note that the following possible outlier values were detected: "1.94", "1.96", "1.97", "1.97", "2.01", "2.02", "2.04", "2.08", "2.14", "2.17" (14 additional values omitted).
- Variable Description: variable measures public debt-to-GDP ratio
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use:
- Issues:

---

tloans

Feature	Result
Variable type	numeric
Number of missing obs.	16693 (88.4 %)
Number of unique values	2169
Median	1187.32
1st and 3rd quartiles	16.1; 25297.5
Min. and max.	0; 311122127.72

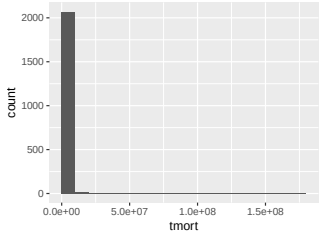


- Note that the following possible outlier values were detected: "654966.88", "656000", "660809", "665687", "670877", "675572", "682748.69", "687835", "695383", "708396" (265 additional values omitted).
- Variable Description: variable measures total loans to non-financial private sector (nominal; local currency)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use: measure to explore link between wealth inequality and loan volume (financial market penetration)
- Issues: needs to be normalized (e.g. by size of economy)

---

tmort

Feature	Result
Variable type	numeric
Number of missing obs.	16770 (88.81 %)
Number of unique values	2072
Median	482.51
1st and 3rd quartiles	5.37; 11588
Min. and max.	0; 178860896

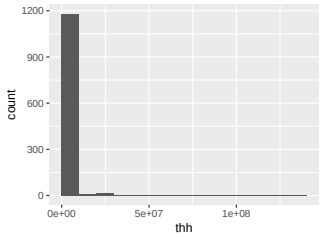


- Note that the following possible outlier values were detected: "311582", "315900", "320443.25", "327615.91", "327661.28", "328837.62", "340395", "350989", "351811", "352246.88" (258 additional values omitted).
- Variable Description: variable measures mortgage loans to non-financial private sector (nominal; local currency)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use: measure to explore link between wealth inequality and loan volume (financial market penetration)
- Issues: needs to be normalized (e.g. by size of economy)

---

thh

Feature	Result
Variable type	numeric
Number of missing obs.	17658 (93.51 %)
Number of unique values	1220
Median	2138.03
1st and 3rd quartiles	152.76; 228262.52
Min. and max.	0.01; 136338661.9

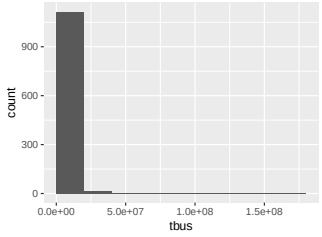


- Note that the following possible outlier values were detected: "6832449.54", "7075941.18", "7537106.92", "8105980.19", "8371255.37", "8637337.35", "8685041.87", "8919444", "9084021.06", "9479231.06" (43 additional values omitted).
- Variable Description: variable measures total loans to households (nominal; local currency)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use: measure to explore link between wealth inequality and loan volume (financial market penetration)
- Issues: needs to be normalized (e.g. by size of economy)

---

tbus

Feature	Result
Variable type	numeric
Number of missing obs.	17733 (93.91 %)
Number of unique values	1146
Median	3030.33
1st and 3rd quartiles	236.85; 224575.13
Min. and max.	0.21; 174783465.81

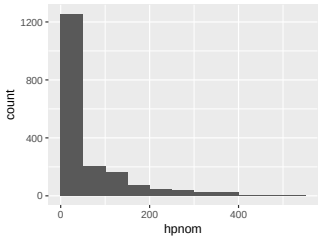


- Note that the following possible outlier values were detected: "6688883.96", "7826015.5", "7861340.94", "9479583.03", "9907019", "11144949.3", "12298568", "12977559.7", "14320429", "14510423.71" (46 additional values omitted).
- Variable Description: variable measures total loans to business (nominal; local currency)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use: measure to explore link between wealth inequality and loan volume (financial market penetration)
- Issues: needs to be normalized (e.g. by size of economy)

---

hpnom

Feature	Result
Variable type	numeric
Number of missing obs.	17047 (90.28 %)
Number of unique values	1737
Median	8.44
1st and 3rd quartiles	2.3; 83.33
Min. and max.	0; 532.09



- Variable Description: variable measures house prices (nominal index; 1990 = 100)
- Source: Jorda-Schularick-Taylor Macrohistory Database, 1870–2015
- How to use:
- Issues:

Report generation information:

- Created by Jonas Markgraf (username: `jonasmarkgraf`).
- Report creation time: Sat Jan 05 2019 21:58:51
- Report was run from directory: `/Users/jonasmarkgraf`
- dataMaid v1.2.0 [Pkg: 2018-10-03 from CRAN (R 3.5.0)]
- R version 3.5.1 (2018-07-02).
- Platform: x86_64-apple-darwin15.6.0 (64-bit)(macOS 10.14.1).
- Function call: `makeDataReport(data = fullData, output = "pdf", useVar = relVars, file = "~/Nexus365/Ben Ansell - WEALTHPOL_RESEARCH/Papers/longrunWealth/code/codebook")`