

# World bank development indicators

Data Incubator Project  
Tim Zeiske  
2016



## World Development indicators

- International financial institution
- Provides loans to developing countries
- Goal: Poverty reduction
- >1000 metrics describing the development of a country/region
- Examples:
  - Arable land (hectares)
  - Net trades (US\$)
  - Electricity from renewables (kWh)
  - CO2 emission (kt)
  - ...

# The data

- CSV file for each country (US, GER)
- Variety of units, scales
- Ranging from 1960 to 2015, values on yearly basis
- Missing datapoints: not all metrics have data for each year

Indicator Code	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
G.AGR.TRAC.NO		4690000	4730000	4755000	4783000	4800000	5470000	5450000	5390000	5305000	5270000	5235000	5205000	5180000	5155000	5130000
G.CON.FERI.PI.ZS																
G.CON.FERI.ZS																
G.LND.AGRI.KZ		4475090	4439910	4414430	4390120	4358730	4355500	4353070	4357950	4354470	4344000	4333000	4323000	4312000	4301580	4291160
G.LND.AGRI.ZS		48.86024177	48.476137029	48.197939504	47.932516356	47.589791854	47.554525841	47.52799444	47.5812756	47.54328002	47.428965734	47.308864762	47.19968206	47.079581088	46.965812712	46.85152424
G.LND.ARBL.HA		180630000	177095000	179574000	177966000	177000000	175705000	174487000	181000000	189244000	188735000	188140000	187545000	187050000	186472000	185879000
G.LND.ARBL.HA.PI		0.983336146	0.9493776067	0.9489119751	0.9274424277	0.9109483642	0.8939000814	0.878089899	0.901816587	0.933722129	0.9204250629	0.9059958297	0.8935139307	0.8826902114	0.8719593742	0.861128047
G.LND.ARBL.ZS		19.72167146	19.335710605	19.606374523	19.430808738	19.325338248	19.183946649	19.05096212	19.76206906	20.66217125	20.606597256	20.541633548	20.47666984	20.422624403	20.359516801	20.29640801
G.LND.CREL.HA		64869558	60548008	62420208	60977008	60379266	61141626	65847777	63042026	59258754	59229150	63782293	58369291	64531443	68262068	72995526
G.LND.CROP.ZS		0.205154297	0.2070104029	0.20406247	0.2044992008	0.2044992008	0.2014420851	0.198712518	0.19598295	0.193253382	0.192707469	0.1921615555	0.191615642	0.1910697284	0.1906329976	0.1901970448
G.LND.ELSM.RU.KZ																
G.LND.ELSM.RU.ZS																
G.LND.ELSM.URKZ																
G.LND.ELSM.UR.ZS																
G.LND.ELSM.ZS																
G.LND.FRST.KZ																
G.LND.FRST.ZS																
G.LND.IRIG.AG.ZS																
G.LND.PRCRMM			715					715					715			
G.LND.TOTL.KZ		9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960
G.LND.TOTL.RU.KZ																
G.LND.TOTL.URKZ																
G.LND.TRAC.ZS		259.6467918	267.08828595	264.79334425	268.7592012	271.18644068	311.31726473	312.3441861	297.7900552	280.3259284	279.22748828	278.25023918	277.53339199	276.93130179	276.44901111	275.96671888
G.PRD.CREL.MI		163619978	162455780	174812487	160937079	183602617	184444880	208158055	202538423	205288170	186860751	237624461	228117866	237683006	204617505	204617505
G.PRD.CROX.XD		43.58	43.98	45.86	44.1	48.16	47.05	49.6	51.62	52.23	49.58	56.77	56.53	60.5	55.16	55.16
G.PRD.FOOD.XD		45.58	45.95	48.23	49.01	50.02	50.98	53.77	54.23	54.34	53.6	58.21	57.38	58.85	60	60
G.PRD.LVSK.XD		55.31	55.48	57.32	60.42	59.43	60.63	62.26	62.79	63.14	64.57	66.06	66.23	63.18	66.46	66.46
G.SRF.TOTL.KZ		9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090
G.YLD.CREL.KG		2522.3	2683.1	2800.6	2639.3	3040.8	3016.7	3161.2	3212.8	3464.3	3154.9	3725.6	3908.2	3683.2	2997.5	2997.5
G.GSR.NFS.VG.D.ZS																
M.GSR.CMCR.ZS																
M.GSR.FET.Y.EB											5070000000	4810000000	5850000000	8950000000	12110000000	12110000000
M.GSR.GNFS.CD											54510000000	61150000000	72850000000	89590000000	125480000000	125480000000
M.GSR.INFS.ZS											39860000000	45580000000	55800000000	70500000000	103820000000	103820000000
M.GSR.MMCH.CD																
M.GSR.NFS.V.CD																
M.GSR.ROFL.CD											2200000000	2400000000	3000000000	4000000000	3600000000	3600000000
M.GSR.TOTL.CD											59580000000	65960000000	78700000000	98540000000	137590000000	137590000000
M.GSR.TRAN.ZS																
M.GSR.TRVL.ZS																
M.KU.DINV.CD.WD																
M.KU.DINV.WD.GD.ZS											64900000000	56200000000	73300000000	93400000000	51600000000	51600000000
M.TREPRV.CD											0.603224883	0.4812591521	0.5715626898	0.6538102648	0.3331557794	0.3331557794
M.TREPWKRC.DI											14800000000	15100000000	15700000000	18300000000	16500000000	16500000000
M.CAB.XOKA.CD											6500000000	7000000000	6700000000	7100000000	6800000000	6800000000
M.CAB.XOKA.GD.ZS											26200000000	-9800000000	-52600000000	75800000000	17000000000	17000000000
M.FIN.TOTL.CD											0.243520677	-0.083920635	-0.41015762	0.5306083306	0.1097606251	0.1097606251
M.GSR.FCT.Y.CD											20700000000	-12728000000	-78790000000	33739000000	-33860000000	-33860000000
M.GSR.GNFS.CD											66800000000	79000000000	89200000000	128300000000	155200000000	155200000000
M.GSR.NFS.CD											-6100000000	-50400000000	-101200000000	-114000000000	-64000000000	-64000000000

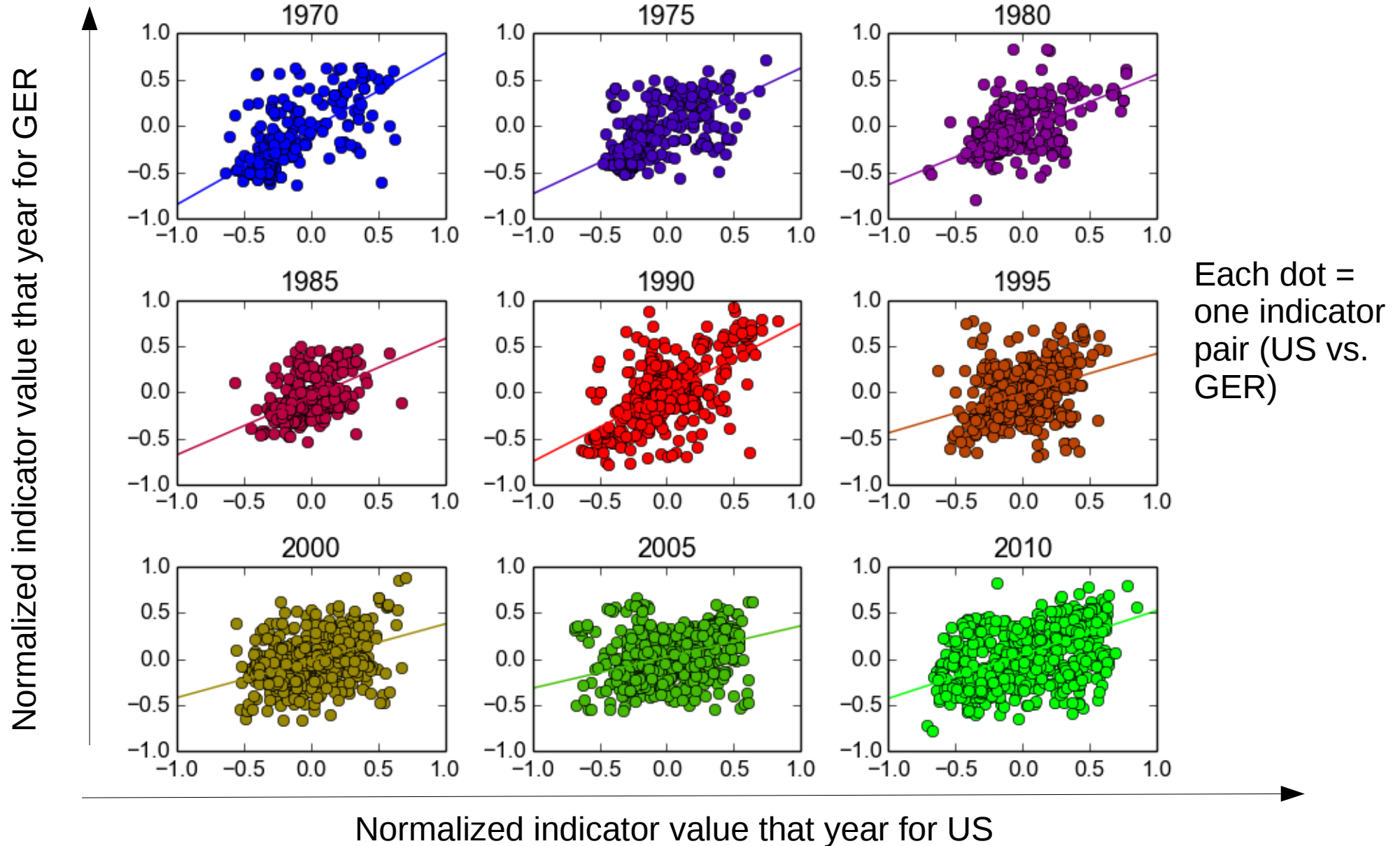
# Quality of the data

- Data decently clean
- Cleanup includes:
  - Identification of data rows/columns, removal or empty ones
  - Removal of missing data points without “shifting” data (data point for 1970 needs to stay assigned to 1970 even if years before are removed)
  - Solution: keep missing data points around as 0.0 and remove only when necessary
- When appropriate, data was scaled by subtracting average for that metric (over all years) and dividing by range (max-min) for that metric

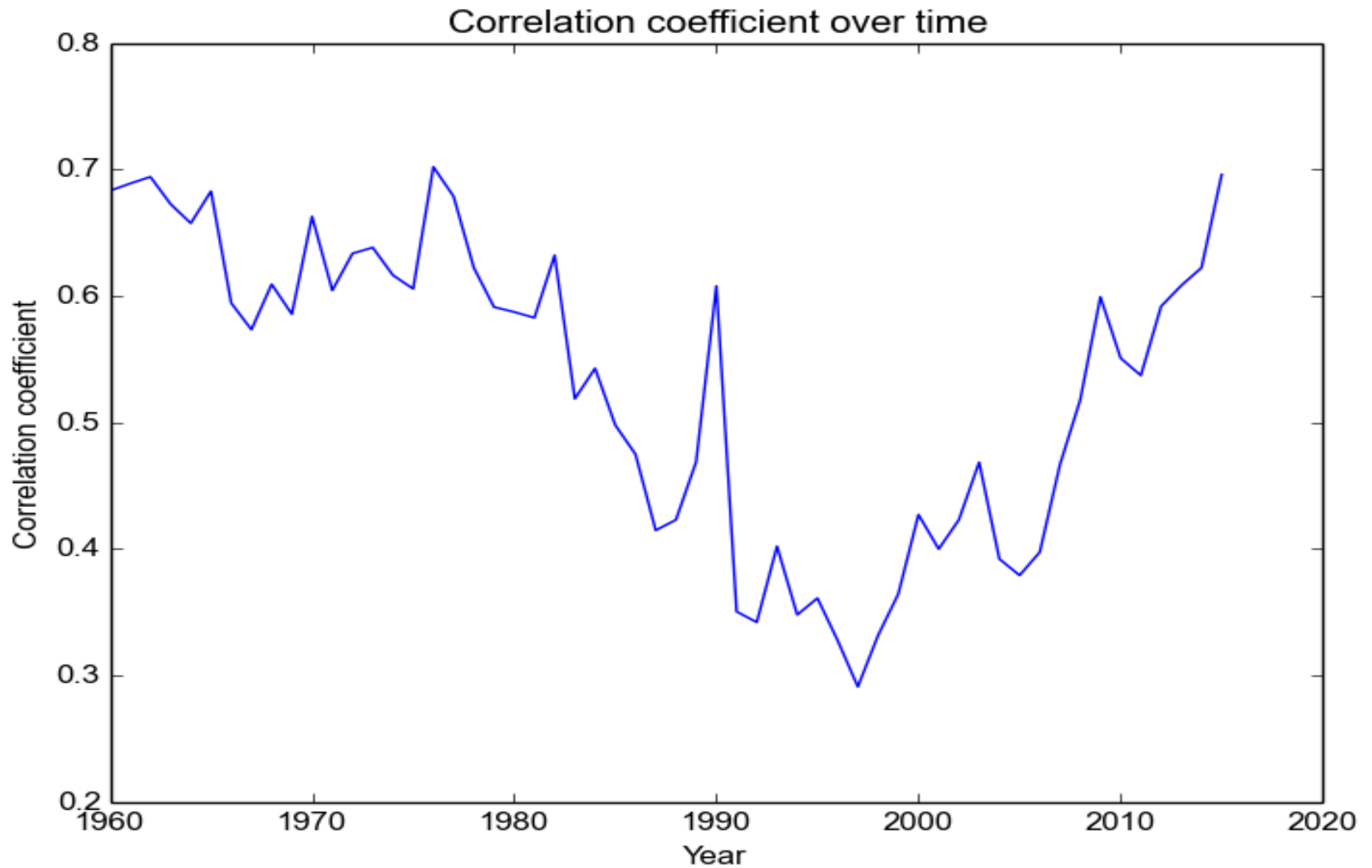
# Tools

- Python
- Numpy
- Matplotlib
- Scipy
- Emacs

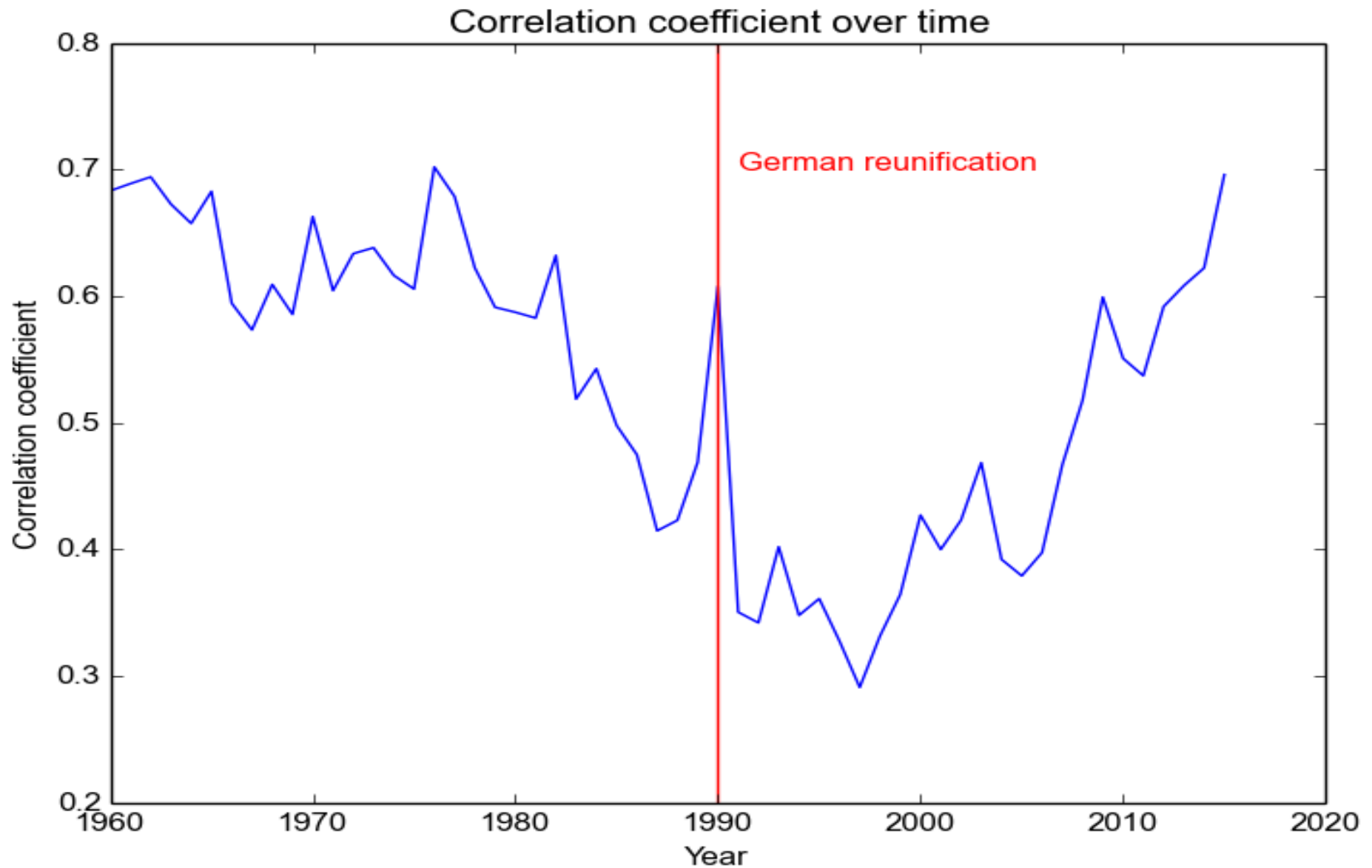
# Correlation between USA & Germany



# Correlation over time

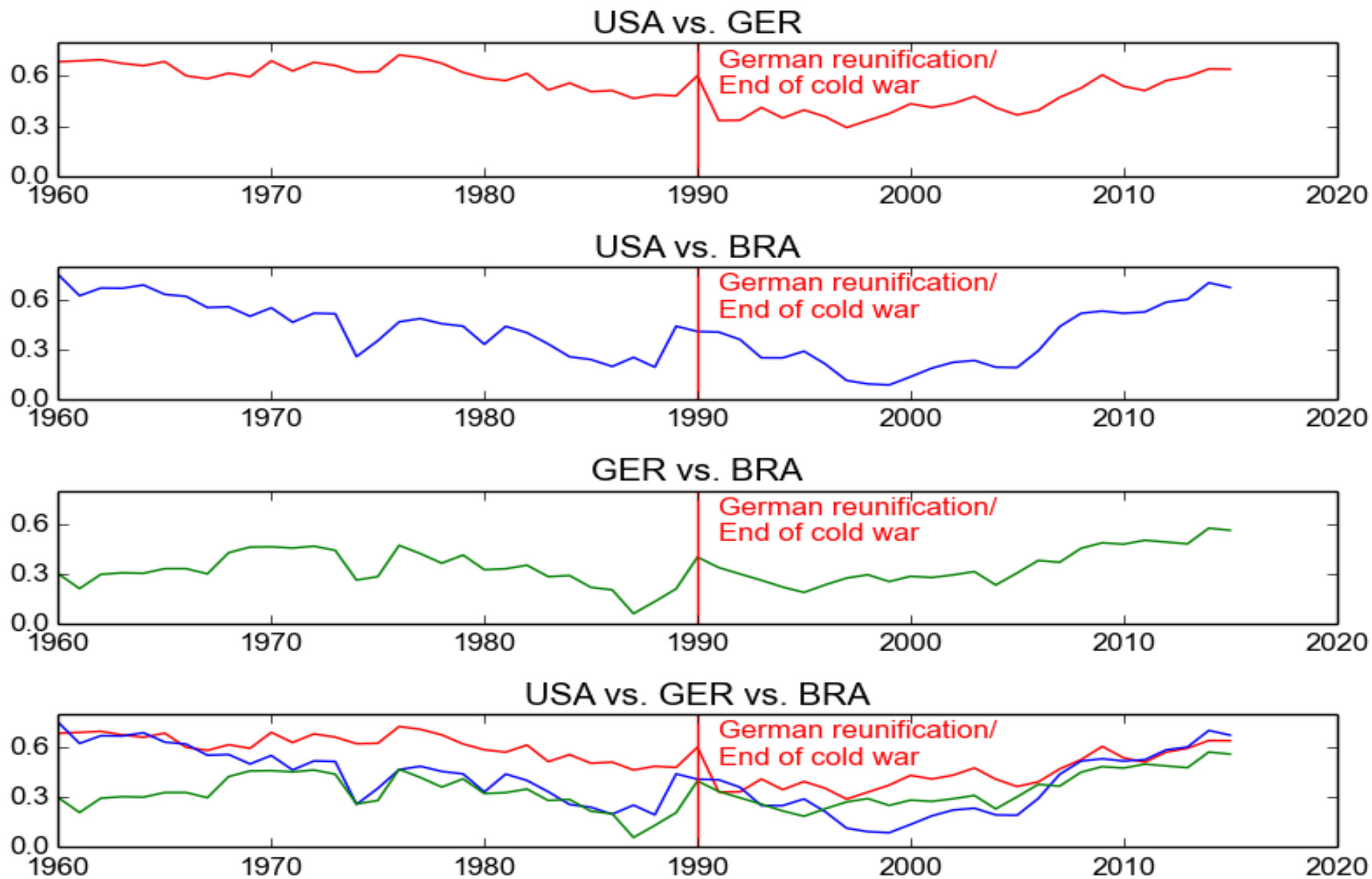


# Historic event: German reunification (end of cold war)

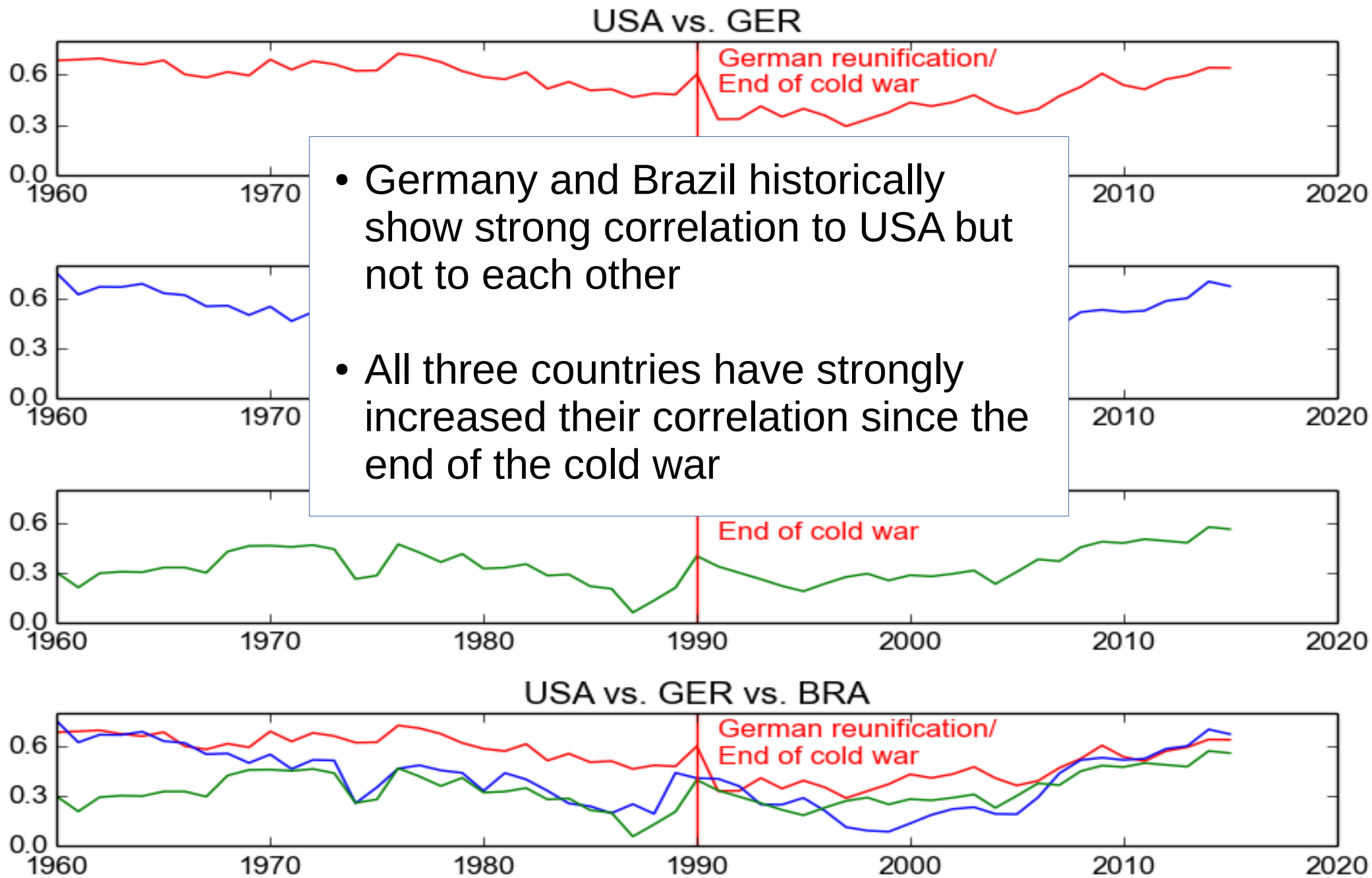




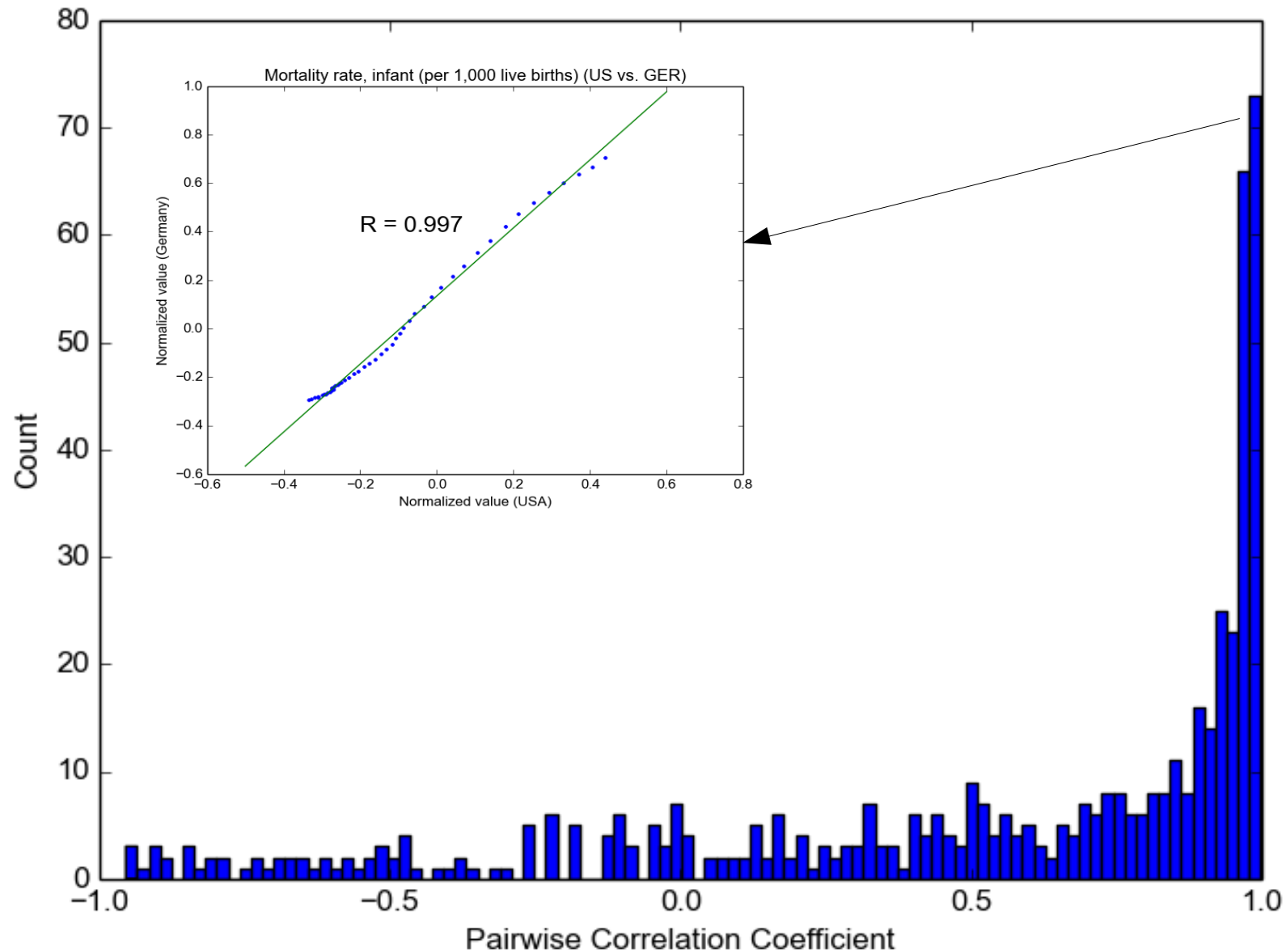
# Third country: Brazil



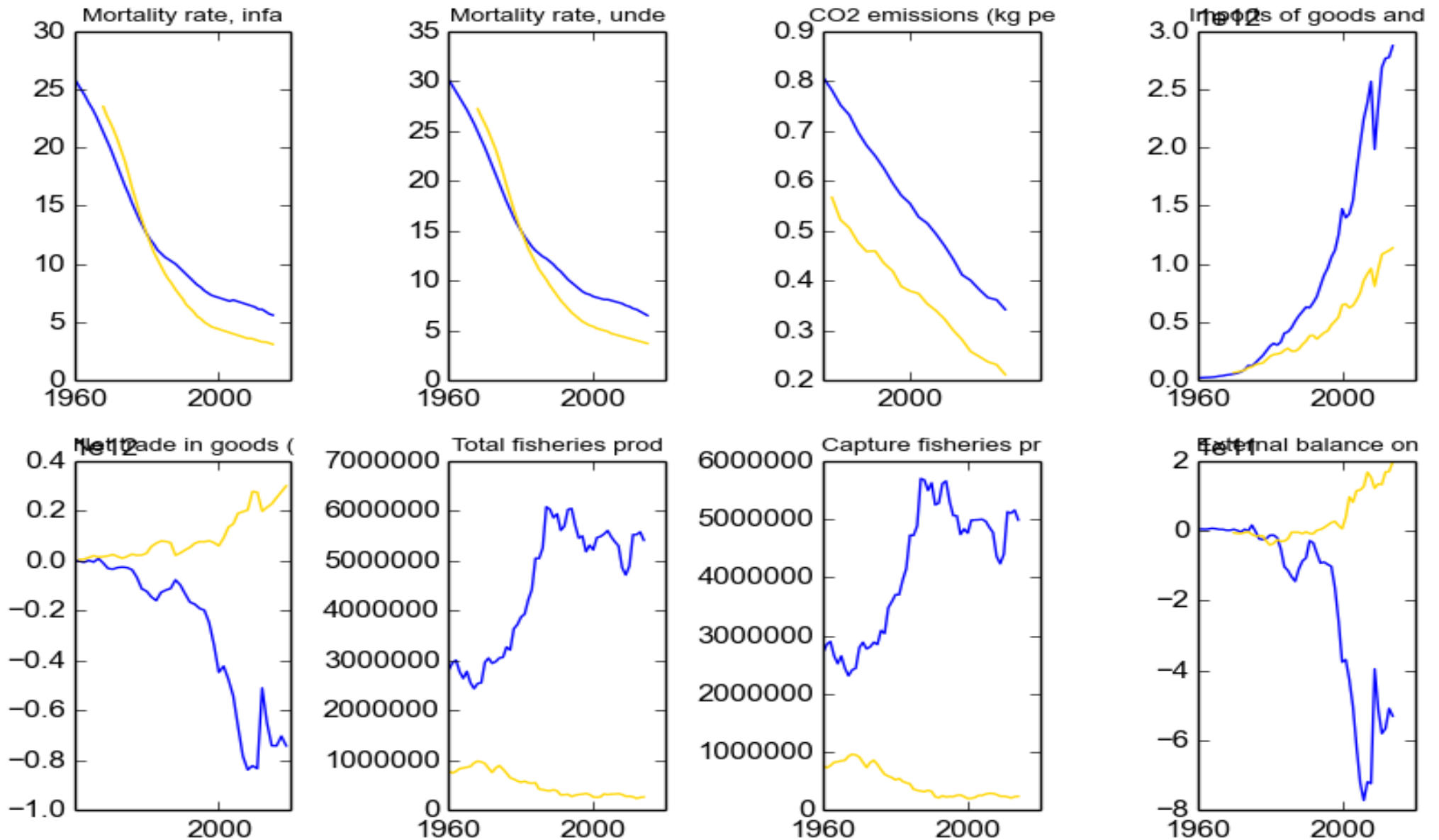
# Third country: Brazil



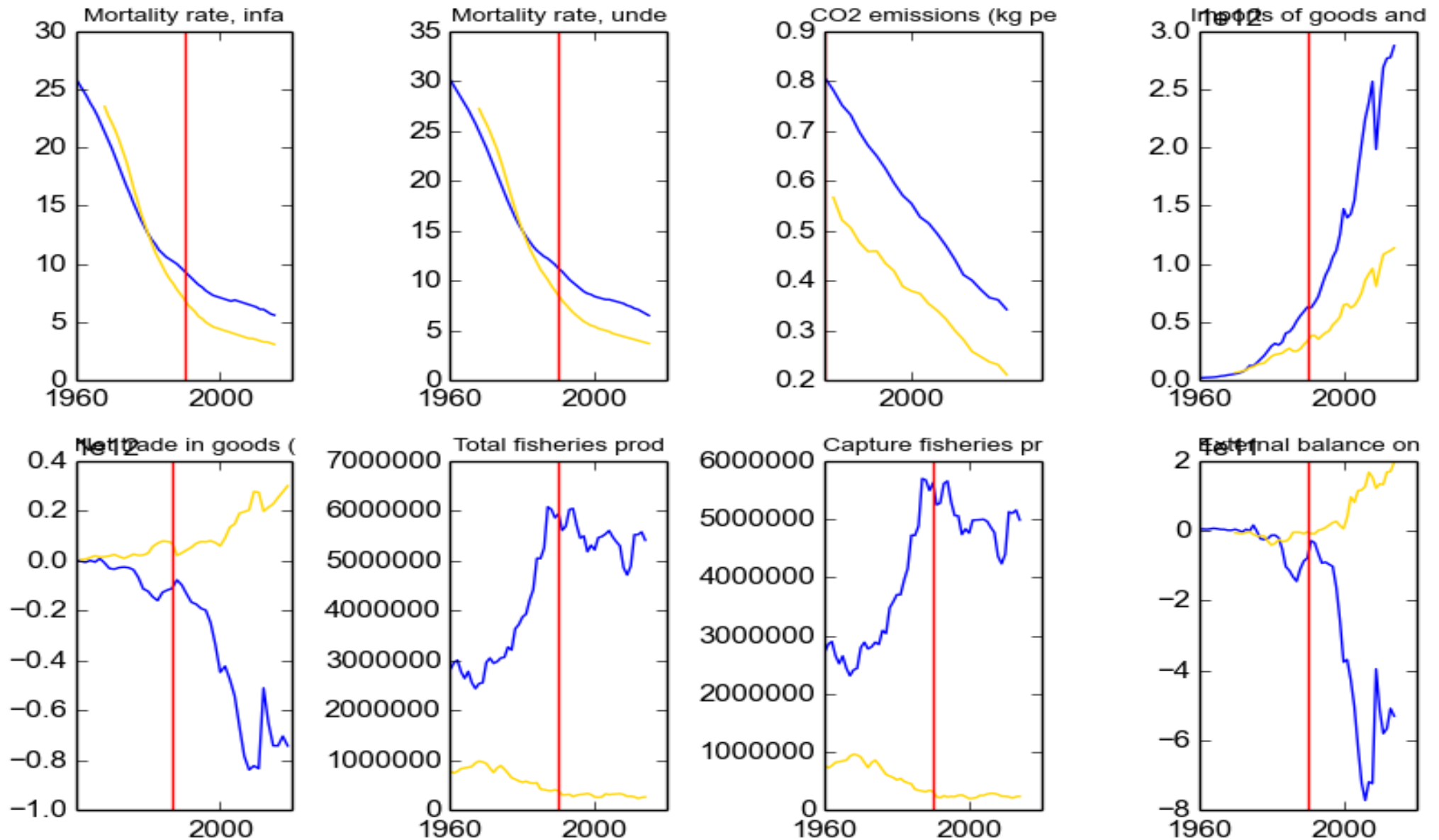
# Correlation coefficients of each indicator between countries (over time)



# Positive and negative correlations



# Influence of reunification



# Future directions

- Correlations between different (not same) indicators within a country/region or across countries?
- Correlations between different indicators across countries and time periods?
- → potential predictive power, for example:
  - Can arable land area in one decade predict CO<sub>2</sub> emissions in the next decade?
  - Will investment in fishery in Germany lead to increased trade balance of USA?