World bank development indicators

Data Incubator Project
Tim Zeiske
2016

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

- 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)

  - CO_2 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

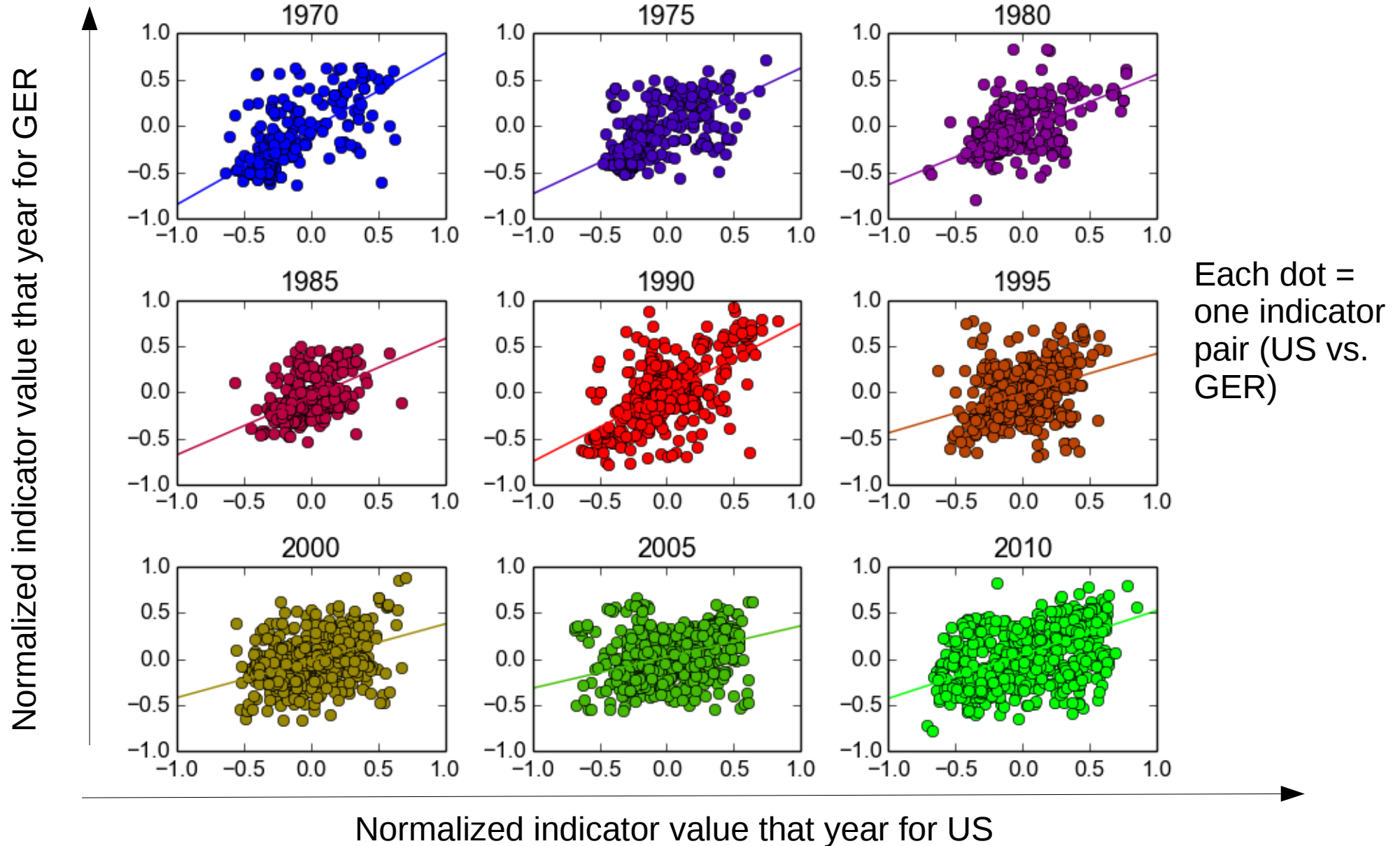| dicator Code	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974
.AGR.TRAC.NO		4690000	4730000	4755000	4783000	4800000	5470000	5450000	5390000	5305000	5270000	5235000	5205000	5180000	5155000
.CON.FERT.PT.ZS															
.CON.FERT.ZS															
.LND.AGRI.K2		4475090	4439910	4414430	4390120	4358730	4355500	4353070	4357950	4354470	4344000	4333000	4323000	4312000	4301580
.LND.AGRI.ZS		48.86024177	48.47613702	48.19793950	47.93251635	47.58979185	47.55452584	47.52799444	47.5812756	47.54328002	47.42896573	47.30886476	47.19968206	47.07958108	46.96581271
.LND.ARBL.HA		18063000	17709500	17957400	17796600	17700000	17570500	17448000	18100000	18924400	18873500	18814000	17545000	18705000	18672000
.LND.ARBL.HA.PC		0.983336146	0.949377606	0.948911975	0.927442427	0.910948364	0.893900081	0.878089899	0.901816587	0.933722129	0.920425062	0.905995829	0.893513930	0.882690211	0.871959374
.LND.ARBL.ZS		19.7217146	19.33571060	19.60634523	19.43080873	19.32533824	19.18394664	19.05096212	19.7620690	20.66217125	20.60659256	20.54163354	20.4766684	20.42262440	20.35951680
.LND.CREL.HA		64869558	60548008	62420208	60977008	60379266	61141626	65847777	63042026	59253754	59229150	63782293	58369291	64531443	68262068
.LND.CROP.ZS		0.205154297	0.207010402	0.2040624	0.204499200	0.204499200	0.201442085	0.19871251	0.19598295	0.193253382	0.192707469	0.192161555	0.191615642	0.191069728	0.190632997
.LND.ELSM.RU.K2															
.LND.ELSM.RU.ZS															
.LND.ELSM.URK2															
.LND.ELSM.UR.ZS															
.LND.ELSM.ZS															
.LND.FRST.K2															
.LND.FRST.ZS															
.LND.IRIG.AG.ZS															
.LND.PRCP.MM			715						715					715	
.LND.TOTL.K2		9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960	9158960
.LND.TOTL.RU.K2															
.LND.TOTL.URK2															
.LND.TRAC.ZS		259.6467918	267.08828595	264.7933425	268.7592012	271.18644068	311.31726473	312.3441861	297.7900552	280.3259284	279.22748828	278.25023918	277.53339199	276.93130179	276.44901111
.PRD.CREL.MT		163619978	162455780	174812487	160937079	183602617	184444880	208158055	202538423	205288170	186860751	237624461	22811/866	237683006	204617505
.PRD.CROP.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
.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
.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
.SRF.TOTL.K2		9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090	9629090
.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
.GSR.NFSV.GD.ZS															
M.GSR.CMCP.ZS															
M.GSR.FCTY.CD											5070000000	4810000000	5850000000	8950000000	12110000000
M.GSR.GNFS.CD											54510000000	61150000000	72850000000	89590000000	125480000001
M.GSR.INSF.ZS															
M.GSR.MRCH.CD											39860000000	45580000000	55800000000	70500000000	103820000000
M.GSR.NFSV.CD															
M.GSR.ROYL.CD											220000000	240000000	300000000	400000000	360000000
M.GSR.TOTL.CD											59580000000	65960000000	78700000000	98540000000	137590000001
M.GSR.TRAN.ZS															
M.GSR.TRVL.ZS															
.KD.DINV.CD.WD											6490000000	5620000000	7330000000	9340000000	5160000000
.KD.DINV.WD.GD.ZS											0.603224883	0.4812591521	0.5715626898	0.6538102648	0.333155779
.TRF.PRVT.CD											1480000000	1510000000	1570000000	1830000000	1650000000
.TRF.PWKR.CD.DT											650000000	700000000	670000000	710000000	680000000
.CAB.XOKA.CD											2620000000	-980000000	-520000000	7580000000	1700000000
.CAB.XOKA.GD.ZS											0.243520677	-0.083920635	-0.410152762	0.530608306	0.109760625
.FIN.TOTL.CD											2070000000	-12728000000	-7879000000	3373900000	-3380000000
.GSR.FCTY.CD											6680000000	7900000000	8920000000	12830000000	15520000000
.GSR.GNFS.CD											-610000000	-5040000000	-10120000000	-11140000000	-6400000000

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

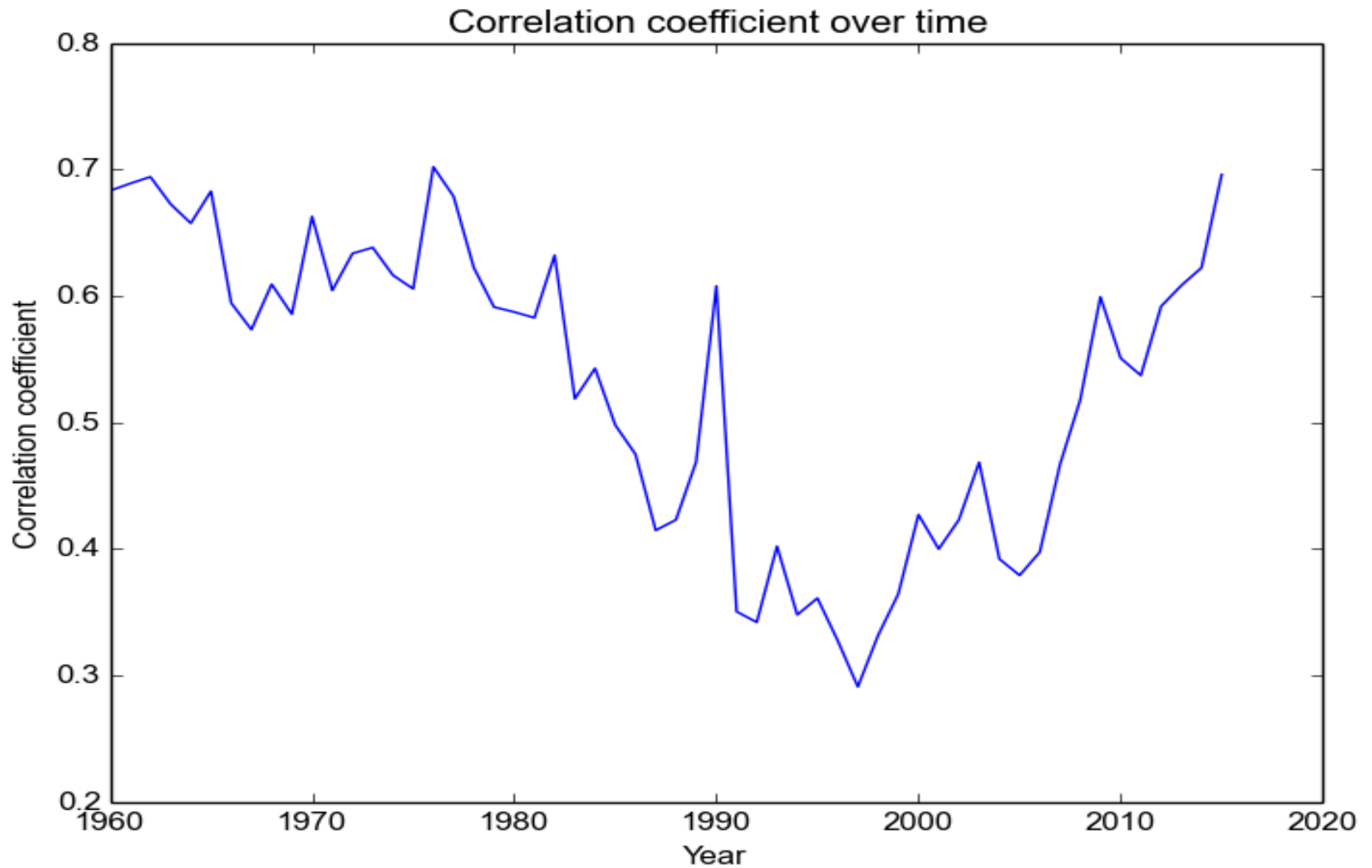- When appropriate, data was scaled by subtracting average for that metric (over all years) and dividing by range (max-min) for that metric
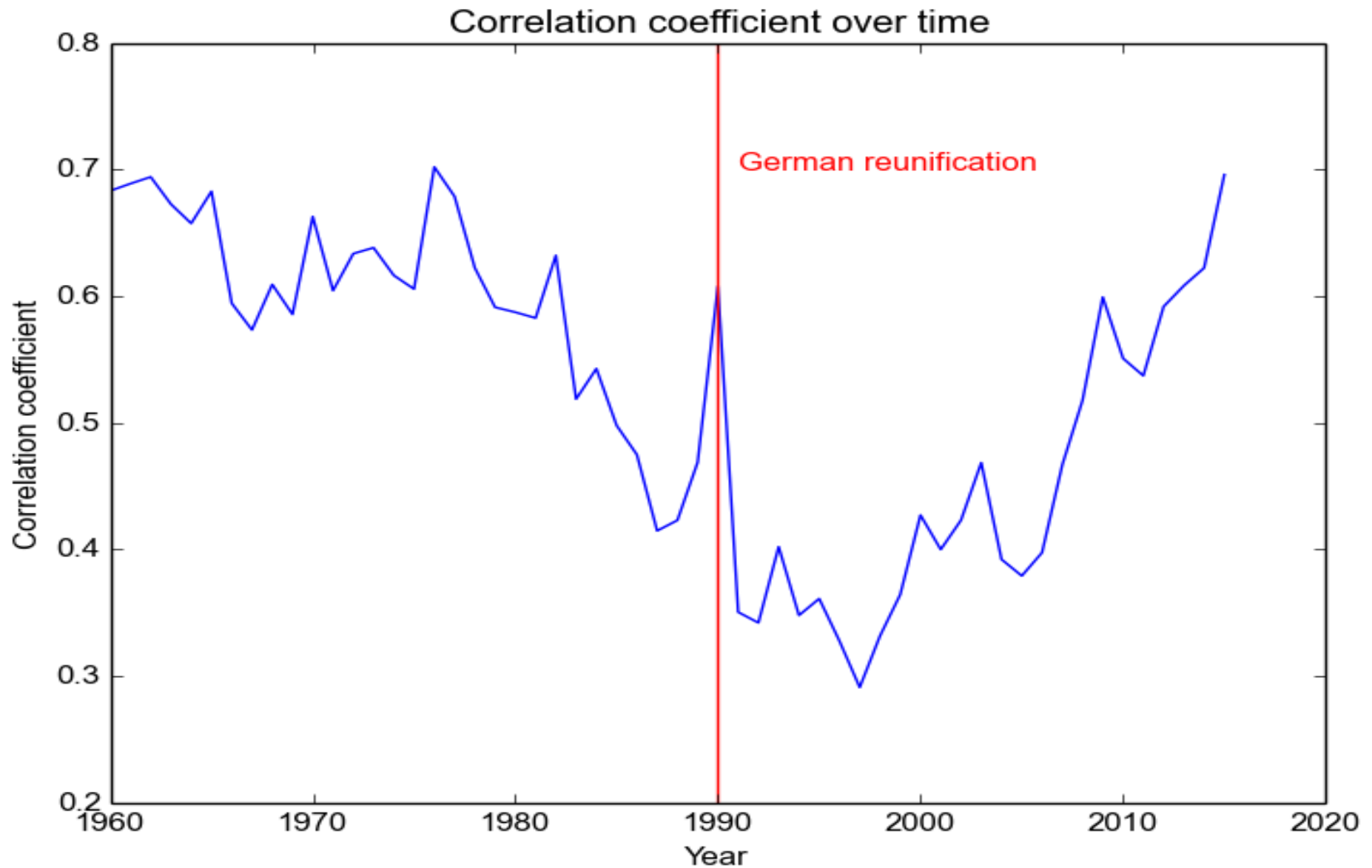
Tools

- Python
- Numpy
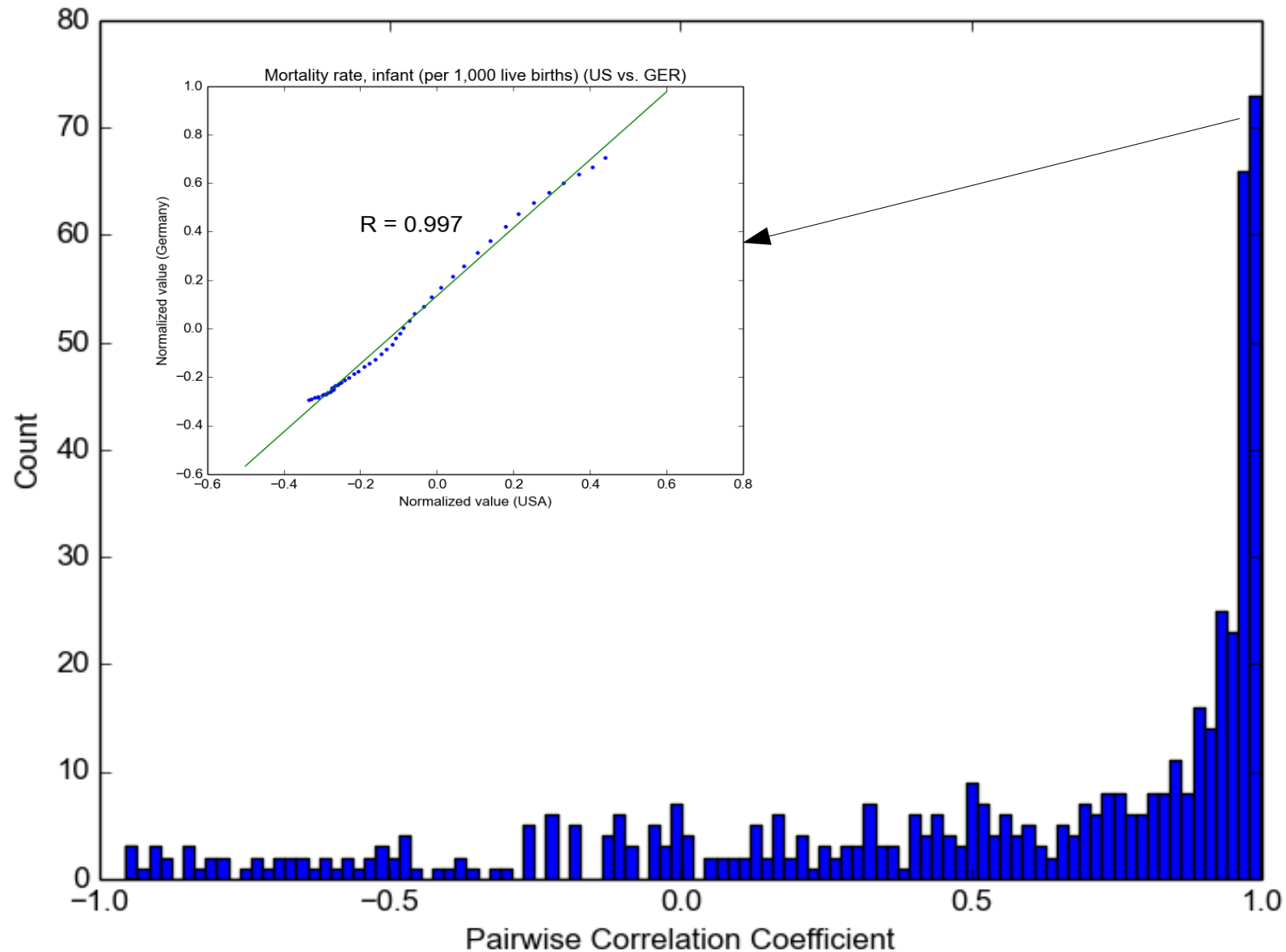- Matplotlib
- Scipy
- Emacs

Correlation between USA & Germany

Each dot = one indicator pair (US vs. GER)

Normalized indicator value that year for GER

Normalized indicator value that year for US
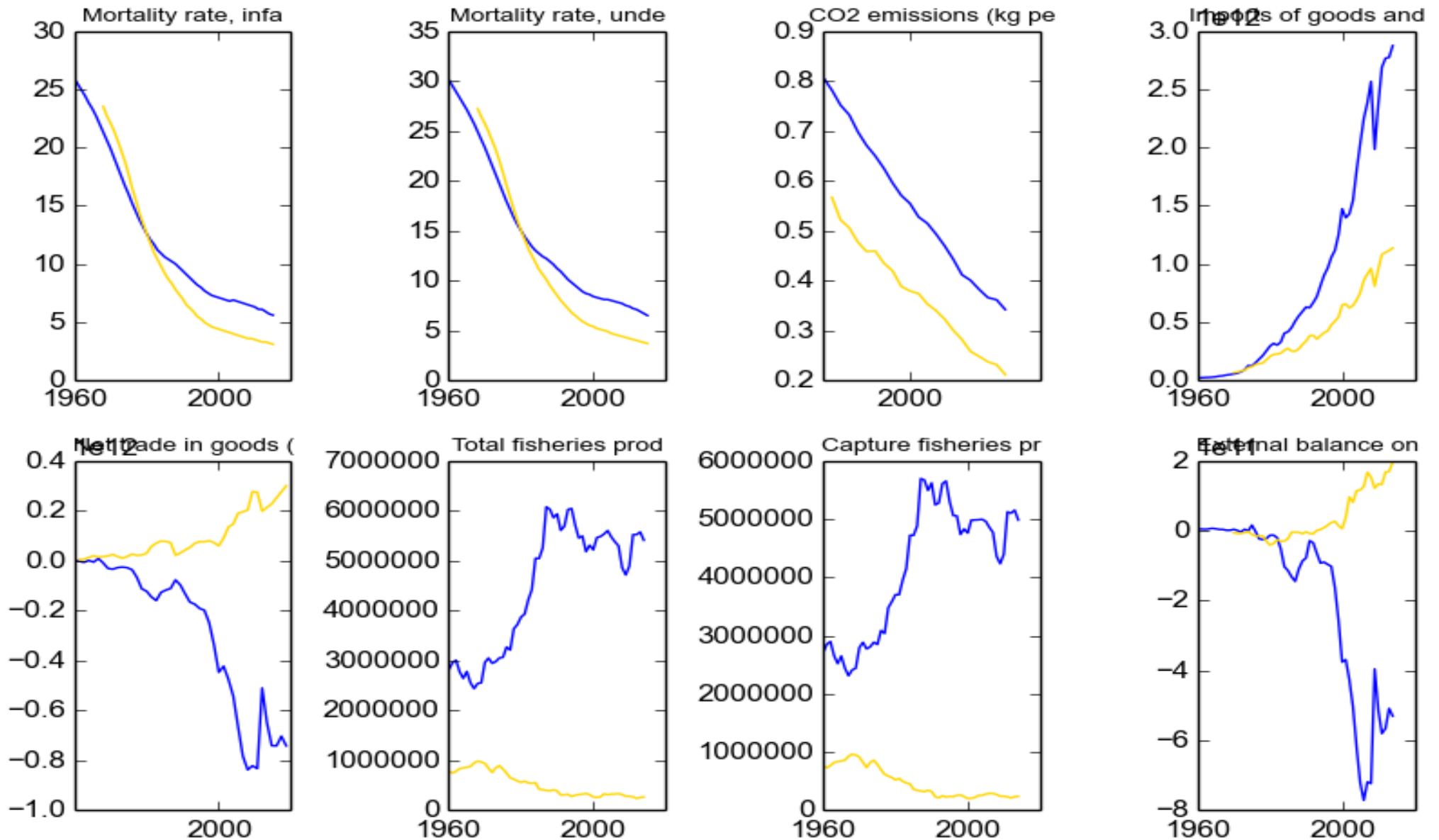
Correlation over time

Historic event: German reunification (end of cold war)
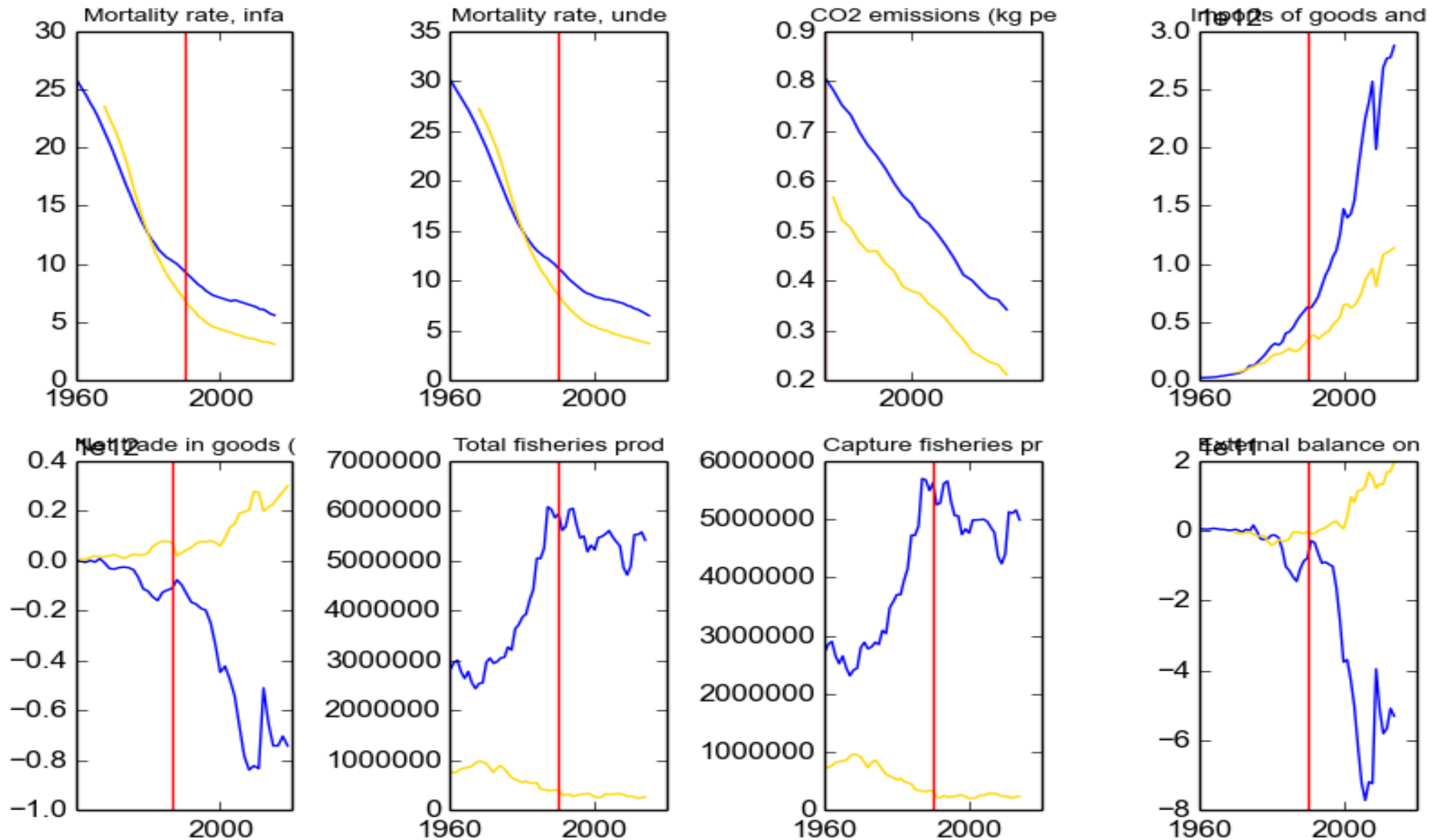
Correlation coefficients of each indicator between countries (over time)

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_2 emissions in the next decade?

  - Will investment in fishery in Germany lead to increased trade balance of USA?