

ขั้นตอนการทำ และปัญหาที่พบเจอ

1. นำข้อมูลเข้า 3 ไฟล์ เป็นไฟล์ .csv โดยเป็นข้อมูล GDP per capita , ราคาตลาดหุ้นของแต่ละประเทศ และชื่อประเทศในแต่ละทวีป
2. มี 2 ไฟล์ที่มีข้อมูลที่ขาดหาย โดยเฉพาะไฟล์ราคาตลาดหุ้น โดยใช้ข้อมูลของปี 2007 - 2022 เนื่องจากบางประเทศตลาดหุ้นเปิดมาได้ไม่นาน หรือบางประเทศมีหลายตลาด และถ้านำข้อมูลจากในเว็บไซด์ทั่วไปจะทำให้ได้ประเทศค่อนข้างน้อย จึงทำการขอข้อมูลจากทางบริษัทให้ช่วยดึงข้อมูลจากโปรแกรม Bloomberg แต่ก็ยังไม่ครบ เนื่องจากบางประเทศไม่มีตลาดหุ้น และบางปีก็ไม่มีข้อมูล เนื่องจากเป็นตลาดใหม่ จึงทำการใส่ค่า 0 ลงไปในช่องว่าง และอีกไฟล์ที่มีข้อมูลว่างคือ GDP per capita เนื่องจากในไฟล์เก็บข้อมูลมาตั้งแต่ปี 1960 ทำให้มีข้อมูลขาดหายไปบ้าง แต่ช่วง 2007-2022 มีบางประเทศที่ไม่มีข้อมูล ก็ทำการใส่ค่าช่องว่างให้เป็น 0 เช่นกัน
3. ชื่อประเทศของแต่ละไฟล์ มีตัวพิมพ์เล็ก และตัวพิมพ์ใหญ่ จึงดำเนินการทำให้ชื่อประเทศ ของทุกไฟล์เป็นตัวพิมพ์เล็ก เนื่องจากจะได้สะดวกต่อการ join ชื่อประเทศ และทวีป
4. ทำการหาเปอร์เซ็นต์การเปลี่ยนแปลงของทั้งตลาดหุ้น และGDPP โดย GDPP ใช้ของปี 2005-2019 ส่วนของตลาดหุ้นใช้ปี 2006-2020 ที่ใช้ปีไม่เท่ากันเนื่องจากว่า GDPP ส่วนใหญ่แต่ละประเทศจะประกาศไตรมาสสุดท้ายของทุกปี ทำให้ตลาดหุ้นมักจะส่งผลหลังจากนั้น ก็คือเล่นในปีถัดไป ทำให้ต้องใช้ข้อมูลของ GDPP ขยับขึ้นมาอีกปี และเมื่อได้ลองหาความสัมพันธ์แล้ว จะมีช่วง COVID-19 ทำให้ข้อมูลผิดปกติไปจากเดิมมาก ทั้งราคาตลาดหุ้น และ GDPP เนื่องจากเศรษฐกิจหยุดชะงักเลยใช้ถึงแค่ปี 2020 เนื่องจาก COVID-19 เกิดช่วงสิ้นปี 2019
5. ได้ทำการจัดรูปแบบของตารางใหม่ ทั้งของGDPP และราคาตลาดหุ้น ให้อยู่ในรูปแบบตารางให้เหมือนกัน เนื่องจากความสะดวกในการหา correlation และการ join ของทั้งปีที่ต้องเหมือนกัน และชื่อประเทศที่ต้องเหมือนกัน
6. ทำ inner join ของทั้ง 3 ไฟล์ โดยแบ่งเป็น 5 ส่วน คือ
 1. GDPP เทียบกับตลาดหุ้นของแต่ละทวีป และแต่ละประเทศ
Inner join ของไฟล์ GDPP ตลาดหุ้น และชื่อประเทศในแต่ละทวีป โดย join ตามปี และชื่อประเทศ
 2. GDPP ของประเทศสหรัฐอเมริกาเทียบกับ GDPP ของแต่ละทวีป และแต่ละประเทศ
Inner join ของไฟล์ GDPP โดยสร้างชื่อตัวแปรใหม่ว่า us เพื่อเก็บข้อมูลของแต่ละประเทศสหรัฐอเมริกา และ join กับประเทศในแต่ละทวีป โดย join ตามปี
 3. GDPP ของประเทศจีนเทียบกับ GDPP ของแต่ละทวีป และแต่ละประเทศ

Inner join ของไฟล์ GDP โดยสร้างชื่อตัวแปรใหม่ว่า china เพื่อเก็บข้อมูลของแต่ละประเทศจีน และ join กับประเทศในแต่ละทวีป โดย join ตามปี

4. ตลาดหุ้นของประเทศสหรัฐอเมริกาเทียบกับ ตลาดหุ้นของแต่ละทวีป และแต่ละประเทศ
Inner join ของไฟล์ตลาดหุ้น โดยสร้างชื่อตัวแปรใหม่ว่า us เพื่อเก็บข้อมูลของแต่ละประเทศสหรัฐอเมริกา และ join กับประเทศในแต่ละทวีป โดย join ตามปี
5. ตลาดหุ้นของประเทศจีนเทียบกับ ตลาดหุ้นของแต่ละทวีป และแต่ละประเทศ
Inner join ของไฟล์ตลาดหุ้น โดยสร้างชื่อตัวแปรใหม่ว่า china เพื่อเก็บข้อมูลของแต่ละประเทศจีนและ join กับประเทศในแต่ละทวีป โดย join ตามปี
7. ทำการหา correlation ของทั้ง 5 ส่วนด้านบนที่กล่าวมา โดยใช้การหาวิธี Spearman's correlation เนื่องจาก Spearman's correlation เป็นการวัดความแรงของความสัมพันธ์ monotonic ระหว่าง 2 ตัวแปร ซึ่งมีค่าตั้งแต่ -1 ถึง 1 โดยที่ข้อมูลไม่ต้องเป็นไปตามเงื่อนไข
 1. ข้อมูลต้องเป็นแบบ interval หรือ ratio level
 2. มีความสัมพันธ์เชิงเส้นตรง
 3. ข้อมูลต้องมีการแจกแจงแบบปกติและเนื่องจากข้อมูลไม่ได้เป็นทั้ง 2 ข้อที่กล่าวมา จึงทำการหาแบบ Spearman's correlation ค่าที่ได้เป็น effect size และสามารถอธิบายความแรงของ correlation ได้ดังนี้
 - .00 - .19 = very weak
 - .20 - .39 = weak
 - .40 - .59 = moderate
 - .60 - .79 = strong
 - .80 - 1.0 = very strong
8. พอได้ค่า Spearman's correlation ออกมา จะนำไป plot heatmap เกิดปัญหาตรงที่กราฟอ่านค่ายาก เนื่องจากข้อมูลในหลายประเทศ และเนื่องจากต้องการหาแค่ความสัมพันธ์ของแต่ละประเทศ ทำให้จึงต้องตัดข้อมูลที่ออกมาจาก Spearman's correlation ไป 1 หลัก และตัดค่าที่เท่ากับ 1 ออก เนื่องจากค่าที่เป็น 1 คือค่าที่ตัวแปรเหมือนกัน มีความสัมพันธ์กันเอง ซึ่งการมีค่านี้นี้ทำให้กราฟอ่านยากขึ้น หลังจากนั้นจึงเรียงข้อมูลจากมากไปน้อย ก็คือที่มีความสัมพันธ์กันมาก จนถึงน้อยและ ไม่มีความสัมพันธ์กัน และเลือกใช้สีแบบ coolwarm เนื่องจากเป็นโทนสีที่ตัดกัน จึงทำให้สามารถมองข้อมูลได้ง่ายขึ้น

สมมติฐาน

1. ตลาดหุ้น และ GDP per capita มีความสัมพันธ์กันไหม
2. ตลาดหุ้นของประเทสหรัฐอเมริกาบ่จีน ใครมีความสัมพันธ์กับประเทศอื่นมากกว่ากัน
3. GDP per capita ของประเทสหรัฐอเมริกาบ่จีน ใครมีความสัมพันธ์กับประเทศอื่นมากกว่ากัน

คำตอบ

1. ตลาดหุ้น และ GDP per capita มีความสัมพันธ์กันไหม

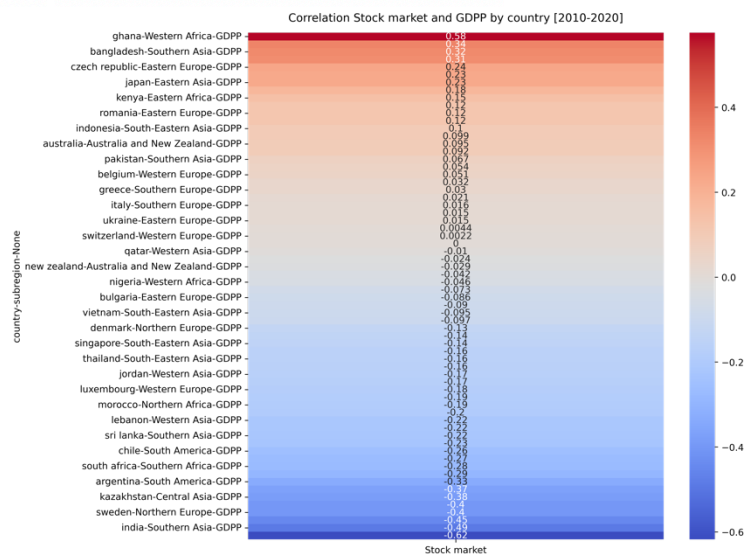
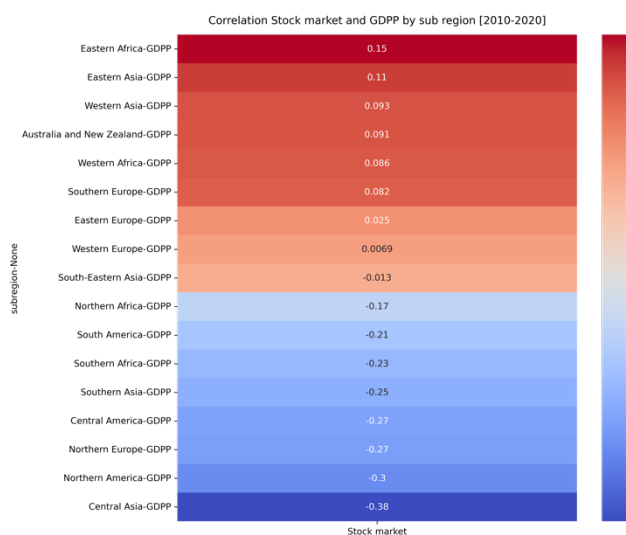
ตอบ แทบจะม่มีความสัมพันธ์กันในข้อมูลของแต่ละทวีป เนื่องจากค่าที่มากที่สุดที่หาแต่ละทวีปได้คือ Eastern Africa (0.15) ซึ่งอยู่ในเกณฑ์ very weak

ส่วนของประเทศ มีประเทศ Ghana (0.58) มีความสัมพันธ์กันระดับ moderate , slovenia และ

Bangladesh (0.34) มีความสัมพันธ์กันระดับ weak และ united arab emirates (0.31) มีความสัมพันธ์กันระดับ weak

Stock market			
country	subregion		
ghana	Western Africa	GDPP	0.575758
slovenia	Southern Europe	GDPP	0.336264
bangladesh	Southern Asia	GDPP	0.321429
united arab emirates	Western Asia	GDPP	0.310393
czech republic	Eastern Europe	GDPP	0.239560

Stock market			
subregion			
Eastern Africa	GDPP	0.145598	
Eastern Asia	GDPP	0.111230	
Western Asia	GDPP	0.093130	
Australia and New Zealand	GDPP	0.090654	
Western Africa	GDPP	0.086087	



2. ตลาดหุ้นของประเทศสหรัฐอเมริกา กับจีน ใครมีความสัมพันธ์กับประเทศอื่นมากกว่ากัน โดยเทียบจากประเทศที่มีค่ามากที่สุด

ตอบ

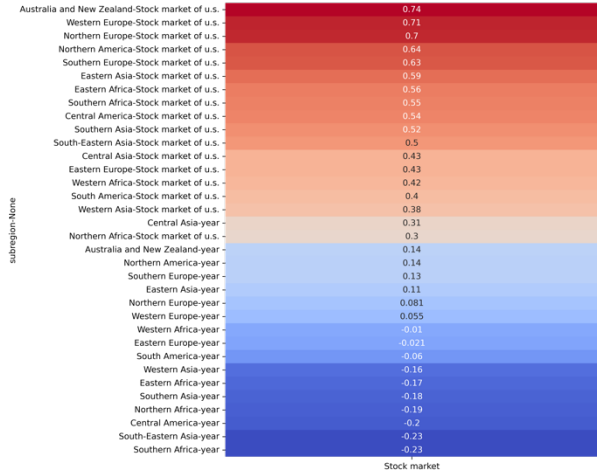
- ถ้าเปรียบเทียบประเทศสหรัฐอเมริกา กับแต่ละทวีป จะมี Australia and New Zealand (0.74) Western Europe (0.71) Northern Europe (0.7) ที่มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ strong
- ถ้าเปรียบเทียบประเทศสหรัฐอเมริกา กับแต่ละประเทศ จะมี switzerland (0.90) sweden และ netherlands (0.80) nigeria (0.8) มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ very strong
- ถ้าเปรียบเทียบประเทศจีน กับแต่ละทวีป จะมี Southern Asia (0.6) Australia and New Zealand (0.6) Eastern Asia (0.58) มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์แบบ strong
- ถ้าเปรียบเทียบประเทศจีน กับแต่ละประเทศ จะมี slovenia (0.75) india (0.74) vietnam (0.72) มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ strong

แต่เมื่อเปรียบเทียบแล้ว ตลาดหุ้นของประเทศสหรัฐอเมริกามีความสัมพันธ์กับแต่ละทวีป และแต่ละประเทศที่สูงกว่า เมื่อเทียบกับประเทศจีน โดยเปรียบเทียบจากประเทศที่มีค่ามากที่สุด

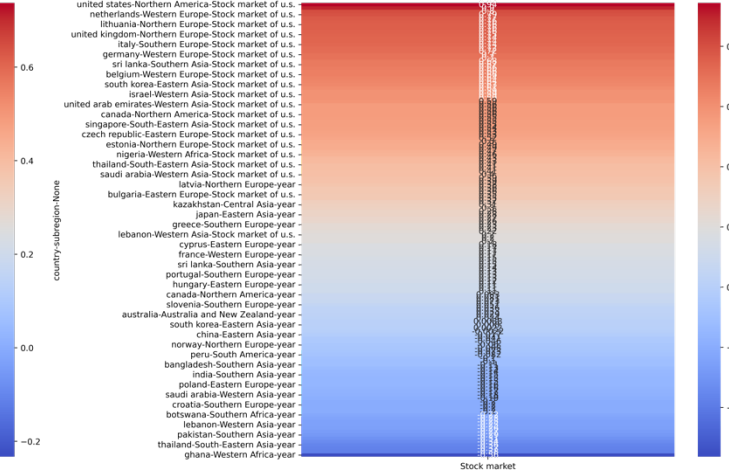
Stock market				Stock market			
subregion				country	subregion		
Australia and New Zealand	Stock market of u.s.	0.737195		united states	Northern America	Stock market of u.s.	0.944171
Western Europe	Stock market of u.s.	0.709850		switzerland	Western Europe	Stock market of u.s.	0.897177
Northern Europe	Stock market of u.s.	0.704891		sweden	Northern Europe	Stock market of u.s.	0.803816
Northern America	Stock market of u.s.	0.636312		netherlands	Western Europe	Stock market of u.s.	0.791735
Southern Europe	Stock market of u.s.	0.628078		norway	Northern Europe	Stock market of u.s.	0.771483

Stock market				Stock market		
subregion				country		
Southern Asia	Stock market of china	0.595829		china	Stock market of china	0.929017
Australia and New Zealand	Stock market of china	0.595623		slovenia	Stock market of china	0.751159
Eastern Asia	Stock market of china	0.575631		india	Stock market of china	0.739906
Southern Europe	Stock market of china	0.566332		vietnam	Stock market of china	0.719078
South-Eastern Asia	Stock market of china	0.561908		israel	Stock market of china	0.704488

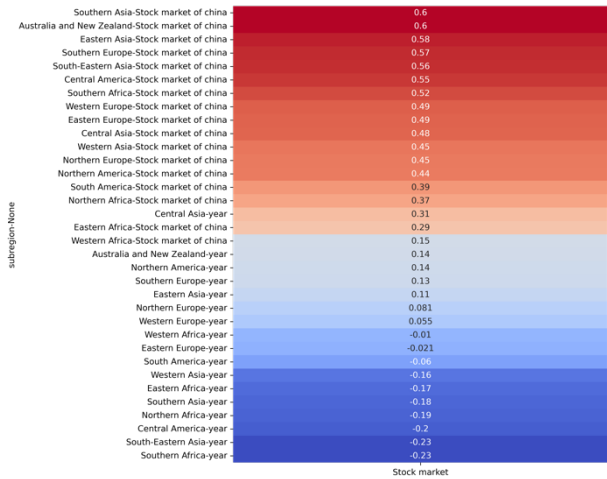
Correlation Stock market of United state and other sub region[2010-2020]



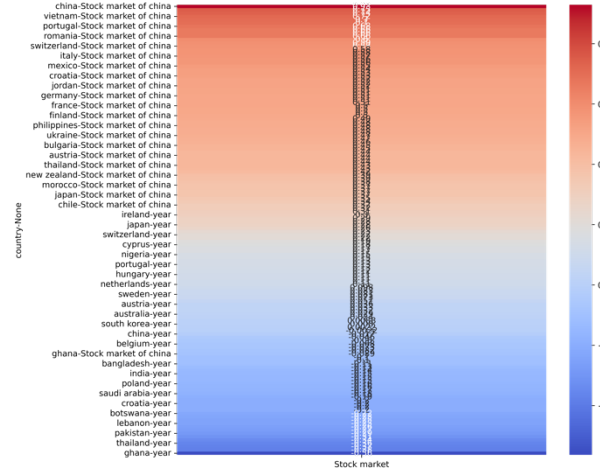
Correlation Stock market of United state and other country[2010-2020]



Correlation Stock market of china and other sub region[2010-2020]



Correlation Stock market of china and other country[2010-2020]



3. GDP per capita ของประเทศสหรัฐอเมริกา กับจีน ใครมีความสัมพันธ์กับประเทศอื่นมากกว่ากัน โดยเทียบจากค่าที่มากที่สุด

ตอบ

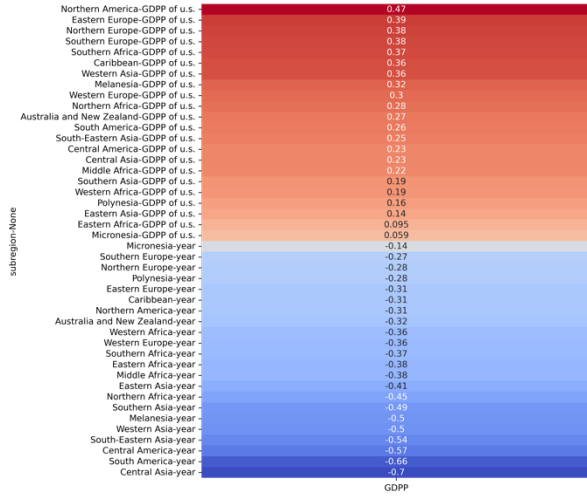
- ถ้าเปรียบเทียบประเทศสหรัฐอเมริกา กับแต่ละทวีป จะมี Northern America (0.47) Eastern Europe (0.39) Northern Europe (0.38) ที่มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ moderate และ weak ตามลำดับ
- ถ้าเปรียบเทียบประเทศสหรัฐอเมริกา กับแต่ละประเทศ จะมี bolivia (0.9) azerbaijan (0.89) antigua and georgia (0.89) มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ very strong
- ถ้าเปรียบเทียบประเทศจีน กับแต่ละทวีป จะมี Central Asia (0.79) Western Asia (0.76) South America (0.75) มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ strong
- ถ้าเปรียบเทียบประเทศจีน กับแต่ละประเทศ จะมี Bolivia (0.9) azerbaijan (0.89) georgia (0.89) มีความสัมพันธ์กันมากที่สุด 3 อันดับแรก โดยมีความสัมพันธ์กันแบบ very strong

แต่เมื่อเปรียบเทียบแล้ว GDP per capita ของประเทศจีน มีความสัมพันธ์กับแต่ละทวีป และแต่ละประเทศที่สูงกว่า เมื่อเทียบกับประเทศสหรัฐอเมริกา เมื่อเปรียบเทียบจากค่าที่มากที่สุด

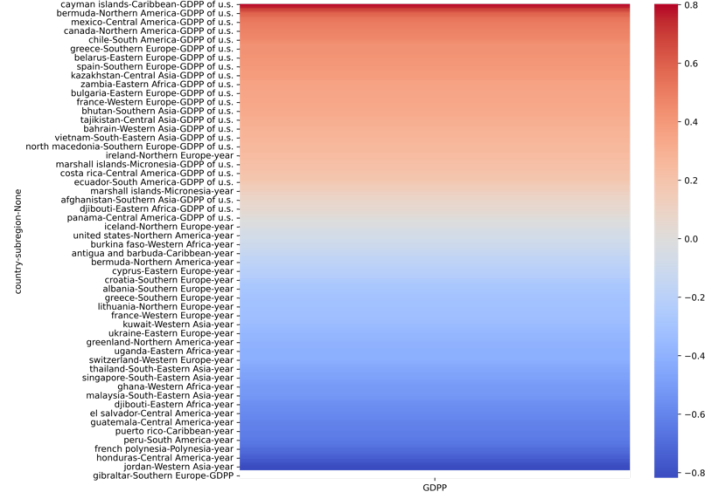
GDPP				GDPP			
subregion				country	subregion		
Northern America	GDPP of u.s.	0.468663		cayman islands	Caribbean	GDPP of u.s.	0.802198
Eastern Europe	GDPP of u.s.	0.394634		curacao	Caribbean	GDPP of u.s.	0.766667
Northern Europe	GDPP of u.s.	0.384342		guinea	Western Africa	GDPP of u.s.	0.760714
Southern Europe	GDPP of u.s.	0.375079		antigua and barbuda	Caribbean	GDPP of u.s.	0.703571
Southern Africa	GDPP of u.s.	0.371633		united kingdom	Northern Europe	GDPP of u.s.	0.671429

GDPP				GDPP			
subregion				country	subregion		
Central Asia	GDPP of china	0.785857		bolivia	South America	GDPP of china	0.900000
Western Asia	GDPP of china	0.759496		azerbaijan	Western Asia	GDPP of china	0.889286
South America	GDPP of china	0.752851		georgia	Western Asia	GDPP of china	0.885714
Melanesia	GDPP of china	0.664150		iraq	Western Asia	GDPP of china	0.885714
South-Eastern Asia	GDPP of china	0.661682		ecuador	South America	GDPP of china	0.871429

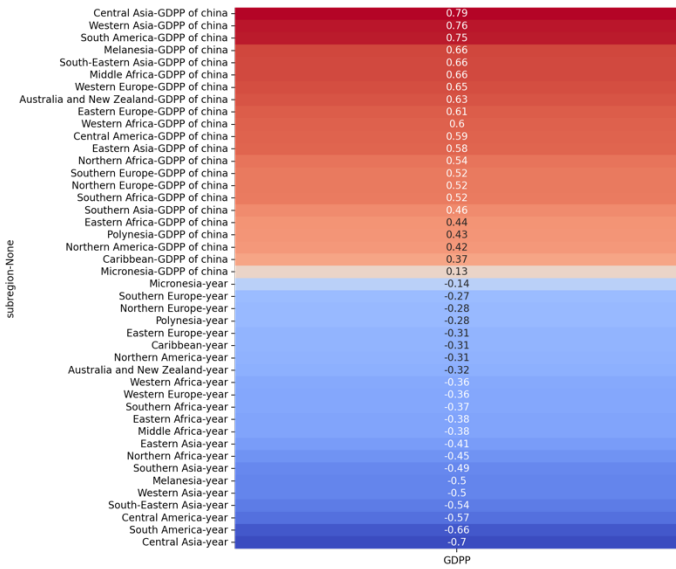
Correlation GDPP of United state and other sub region[2010-2020]



Correlation GDPP of United state and other country[2010-2020]



Correlation GDPP of China and other sub region[2010-2020]



Correlation GDPP of China and other country[2010-2020]

