

Hiddenを用いたデータ解析

<テーマ>

新型コロナウイルスの各国の感染率・死亡率の違いについて

設定した軸

コロナウイルス感染率(%)・・・感染者÷人口×100

コロナウイルス死亡率(%)・・・死亡者数÷人口×100

人口(万人)

平均寿命

緯度

経度

GDP(百万US \$)

出生数(1人の女性が出産する子供数の平均)

死亡率(単位期間あたりの1000人に対する死亡者数)

幸福度・・・世界幸福度調査(World Happiness Report 2020より)

貧困率・・・その国で貧困層と定義される人口の割合(%)

医師の人数・・・人口1000人あたり(人)

PM10($\mu\text{g}/\text{m}^3$)

供給カロリー・・・1人1日当たり(kcal)

飲酒量・・・1人あたりアルコール消費量(L)

喫煙率(%)

喘息死亡率・・・10万人当たり(人)

※値が見つからなかった所や極端に大きい点は数値を操作しています。

データ引用元

<https://www.globalnote.jp/post-1555.html>

<https://datatopics.worldbank.org/world-development-indicators/>

(世界の国別人口)

<https://graphics.reuters.com/CHINA-HEALTH-MAP-LJA/0100B5FZ3S1/index.html>

(コロナ感染者数・死亡者数)

https://memorva.jp/ranking/unfpa/who_whs_life_expectancy.php

(平均寿命)

<http://www.グーグルアース4.jp/data/list00.html>

(世界の緯度・経度)

<https://www.globalnote.jp/post-1409.html>

(国別GDP)

<http://top10.sakura.ne.jp/WHO-WHS9-95.html>

(国別出生数)

<http://top10.sakura.ne.jp/CIA-RANK2066R.html>

(国別死亡率)

データ引用元

<https://happiness-report.s3.amazonaws.com/2020/WHR20.pdf>

(世界幸福度調査)

<http://top10.sakura.ne.jp/CIA-RANK2046R.html>

(国別貧困率)

<http://top10.sakura.ne.jp/IBRD-SH-MED-PHYS-ZS.html>

(国別医師数)

<http://top10.sakura.ne.jp/IBRD-EN-ATM-PM10-MC-M3.html>

(PM10)

<http://honkawa2.sakura.ne.jp/0100.html>

(国別供給カロリー)

<http://top10.sakura.ne.jp/WHO-SA01400-SA-TOTAL.html>

(国別アルコール消費量)

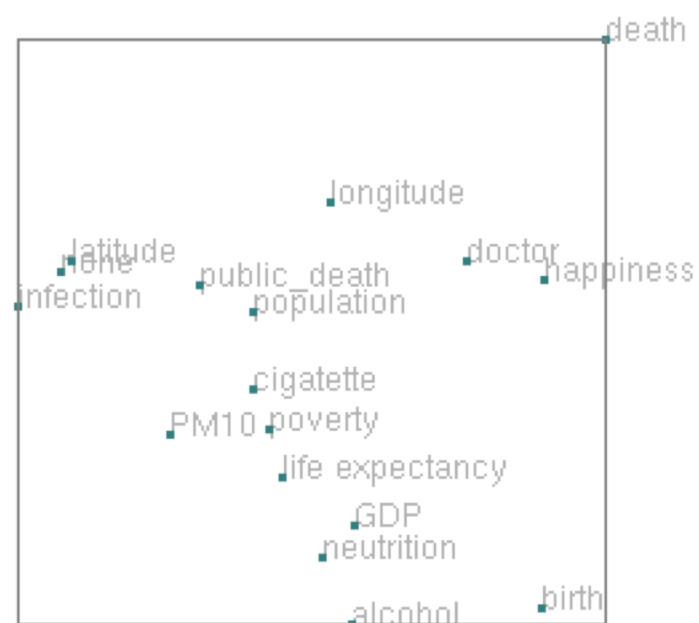
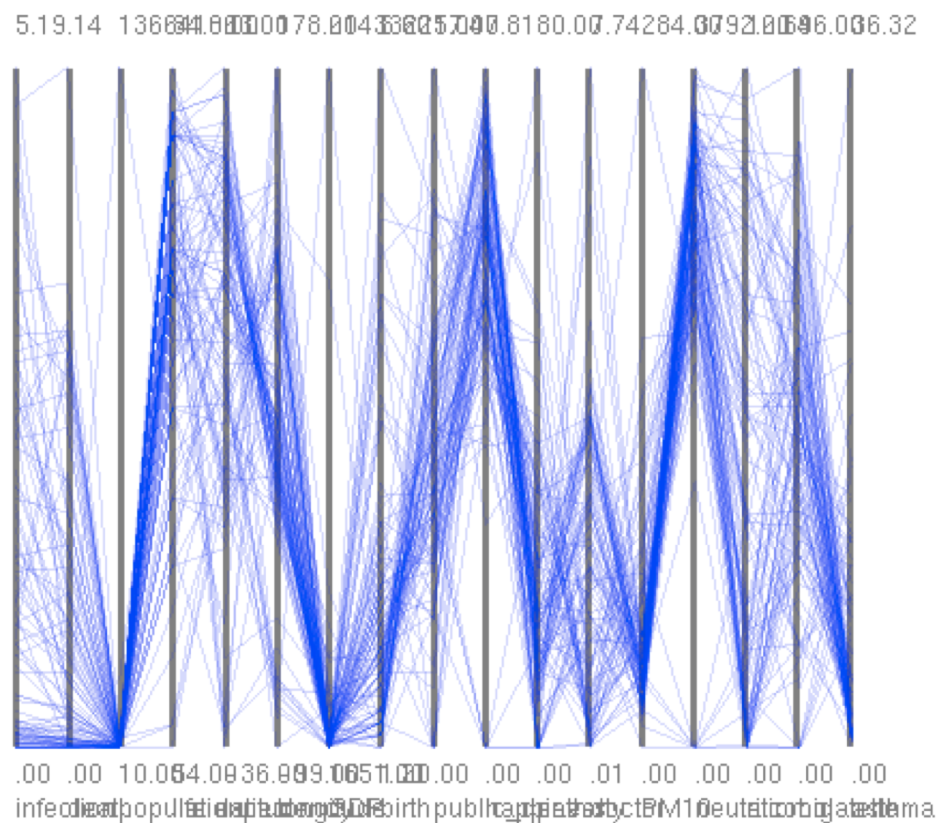
<http://top10.sakura.ne.jp/WHO-M-Est-cig-curr.html>

(喫煙率)

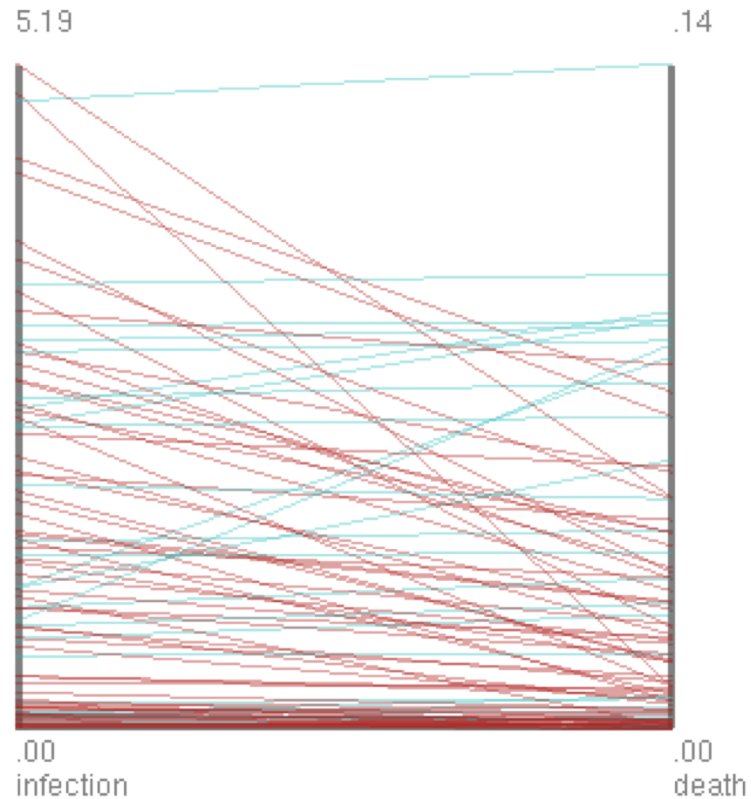
<http://top10.sakura.ne.jp/WHO-GHE119.html>

(喘息死亡者数)

実行結果

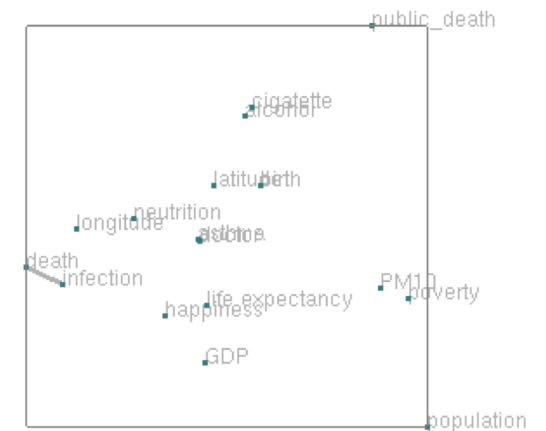


考察(コロナ感染率とコロナ死亡率の関係)

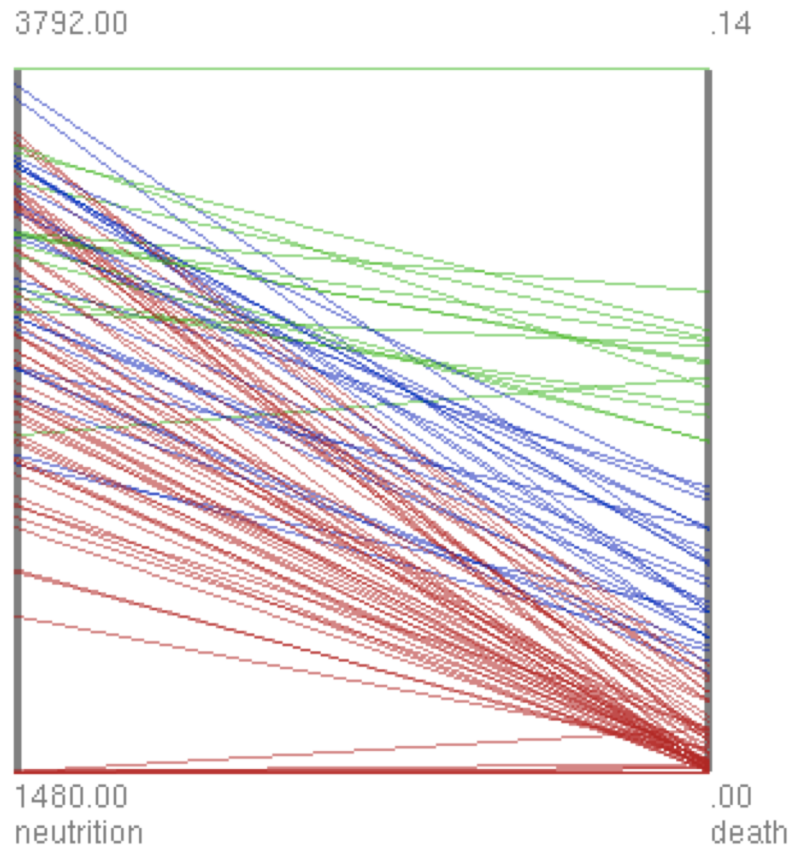


正の相関がある

→感染者数が多い国ほど死者数も多い。この2つの相関は他の2点の相関より高いため、散布図では一番最初に繋がった。



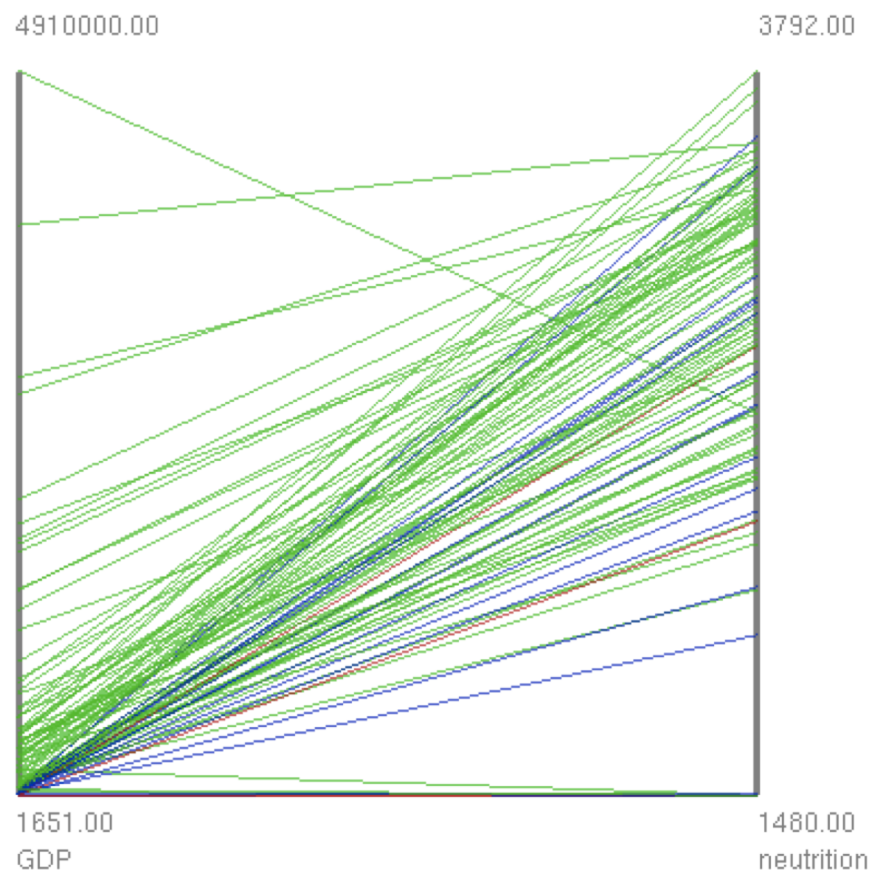
考察(コロナ死亡率と1日の供給カロリーの関係)



正の相関がある

→1日の供給カロリーが高いほどコロナ死亡率が高い。

考察(1日の供給カロリーとGDPの関係)

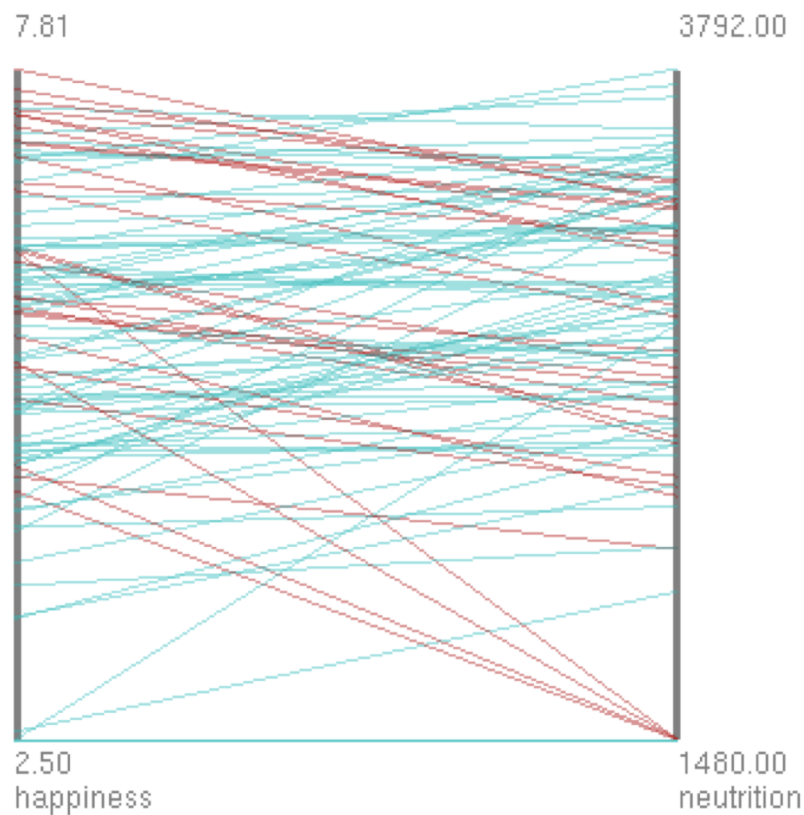


正の相関がある

→GDPが高い国、すなわち先進国(のことが多い)ほど1日の供給カロリーが高い。

先進国ほど1人あたりの生活に余裕があり、多くの食料を手に入れられることが原因と考えられる。

考察(1日の供給カロリーと幸福度の関係)



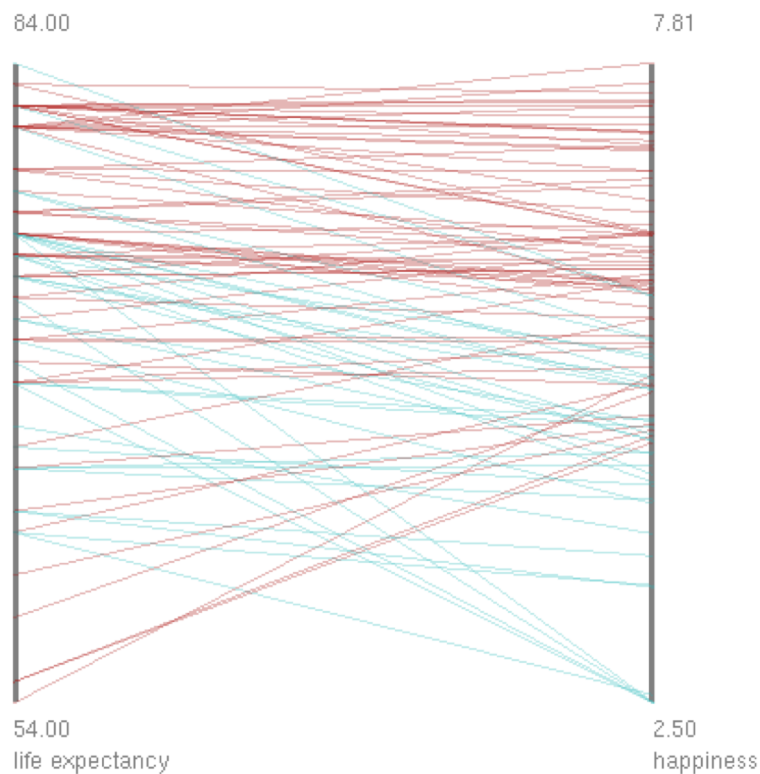
正の相関がある

傾きが正の直線(青が多い)と傾きが負の直線が引かれているが、それぞれに注目すると幸福度が高い国ほど供給カロリーが高い傾向は保たれている

→幸福度が高い国ほど1日の供給カロリーが高い。

食欲が満たされることが個人が感じる幸福度に良い影響を与えることが想像できる。

考察(幸福度と平均寿命の関係)

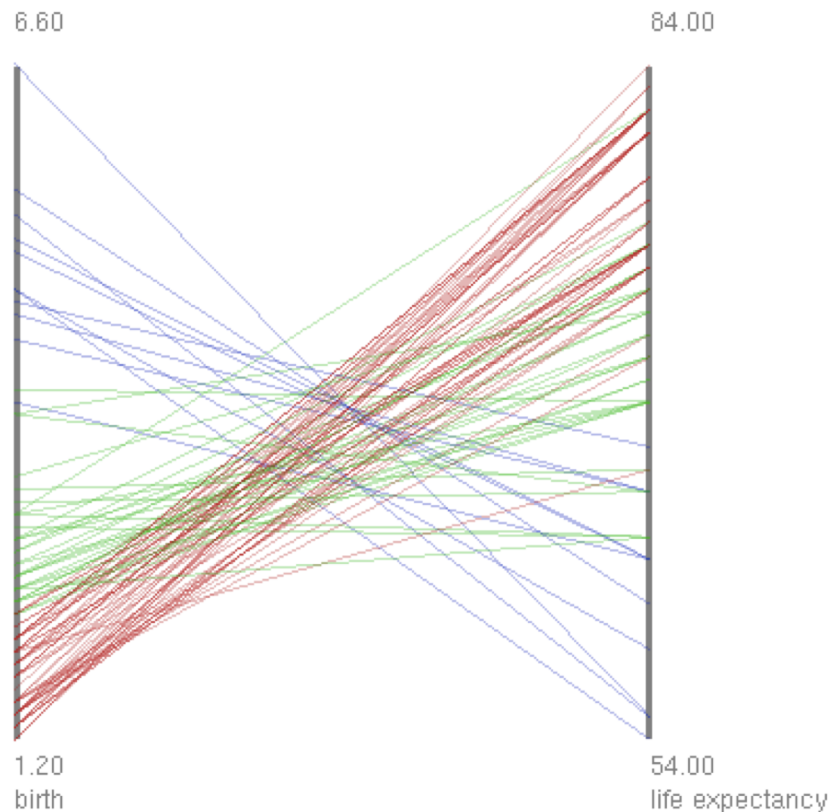


正の相関がある

→幸福度が高い国ほど平均寿命が高い。
平均寿命が同じ国でも幸福度の高さに差があるが、
幸福度が低い(青)国の中でも平均寿命が長い国の方が幸福度が
高い傾向は保たれている。先進国ほど幸福度が高いことが前の
スライドからわかるので、先進国ほど平均寿命が高いことがわ
かる。

幸福度を測る基準の中に平均寿命も含まれていることが原因と
考えられる。

考察(平均寿命と出生数の関係)



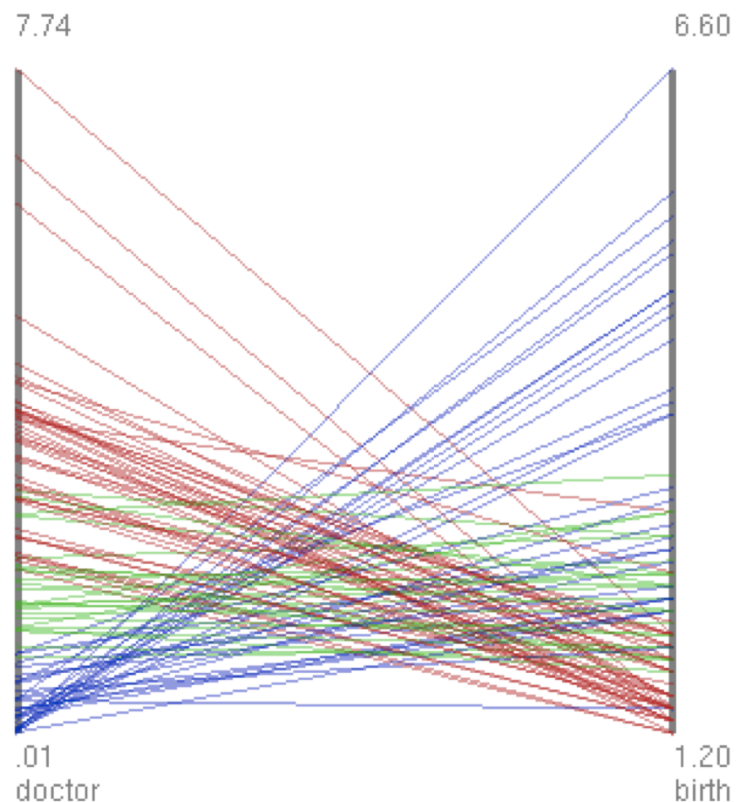
全体的には負の相関がある

出生数が2人以下の国、2人から4人、4人から6.6人に分けるとそれぞれの中で正の相関がある

→出生数が低い国、中くらいの国、低い国の順で平均寿命は低くなっていく。平均寿命が高い国は先進国であることがわかっているので、赤が先進国を表していて、先進国の方が出生数が低い傾向があることがわかる。

発展途上国では1世帯あたりの子供の人数が多く、一方先進国では女性も働き、子供を出産しない家庭も多くあることから推測できる。

考察(出生数と医師の人数の関係)

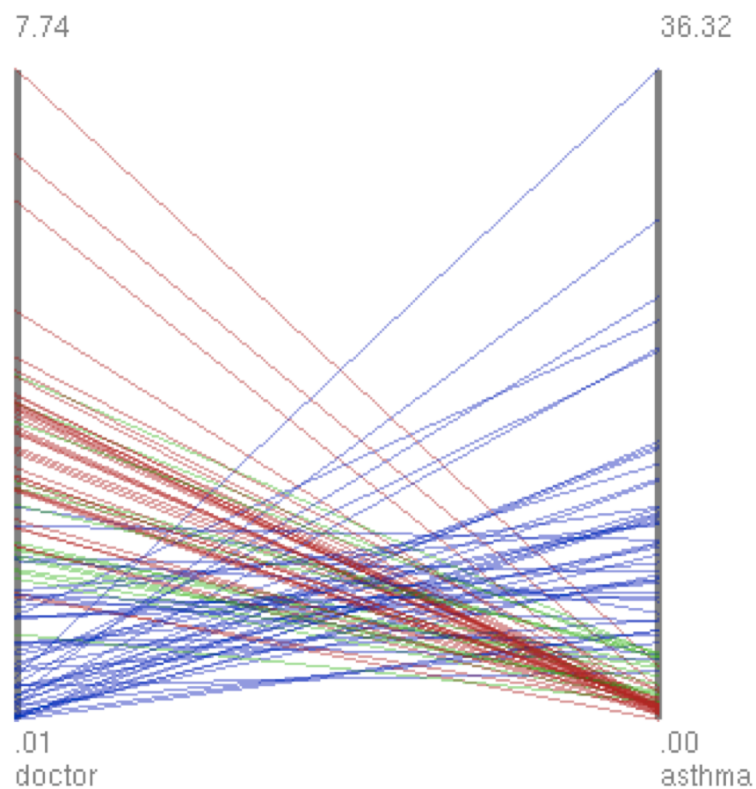


全体的には負の相関がある
出生数が2人以下の国、2人から4人、4人から6.6人に分けると3つの中心で交差している

→出生数が高い国ほど医師の数は少ない。前のスライドで、出生数が高い国は平均寿命が低く発展途上国であることがわかったことから、発展途上国ほど医師の数が少ないことがわかる。

医師になるには多くの学費もかかるため、先進国の方が大幅に医師の数が増えることが想像できる。

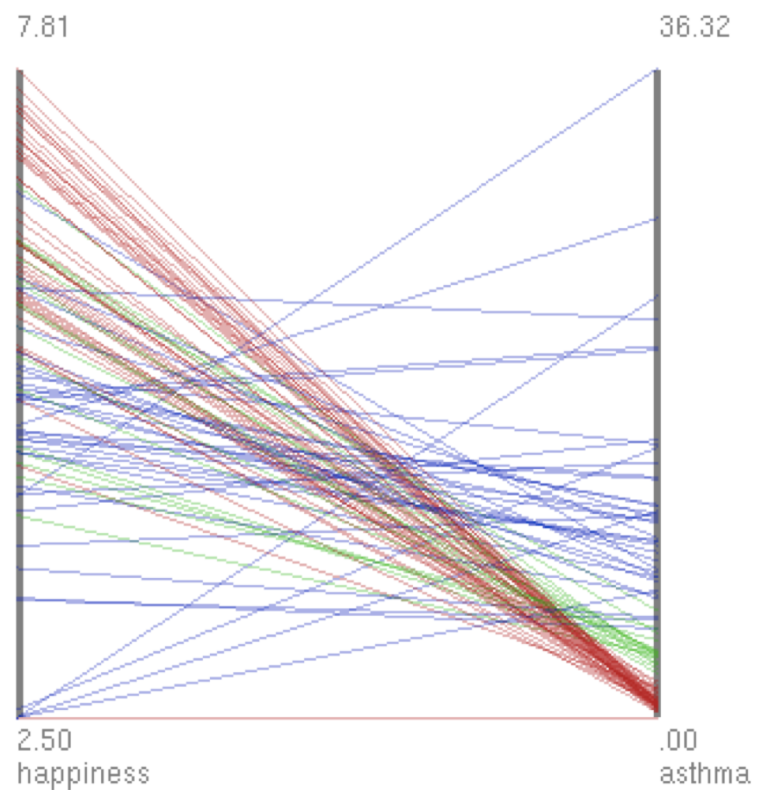
考察(医師の数と喘息死亡者数の関係)



全体的には負の相関がある
医者数が多い、中くらい、少ないところで分けると
それぞれの中で正の相関がある

→単位人数あたりの医者数が多いほど喘息での死者が
少ない。前のスライドで医者数が多いほど先進国であることが
わかったので先進国ほど喘息での死者数が少ないことがわかる。

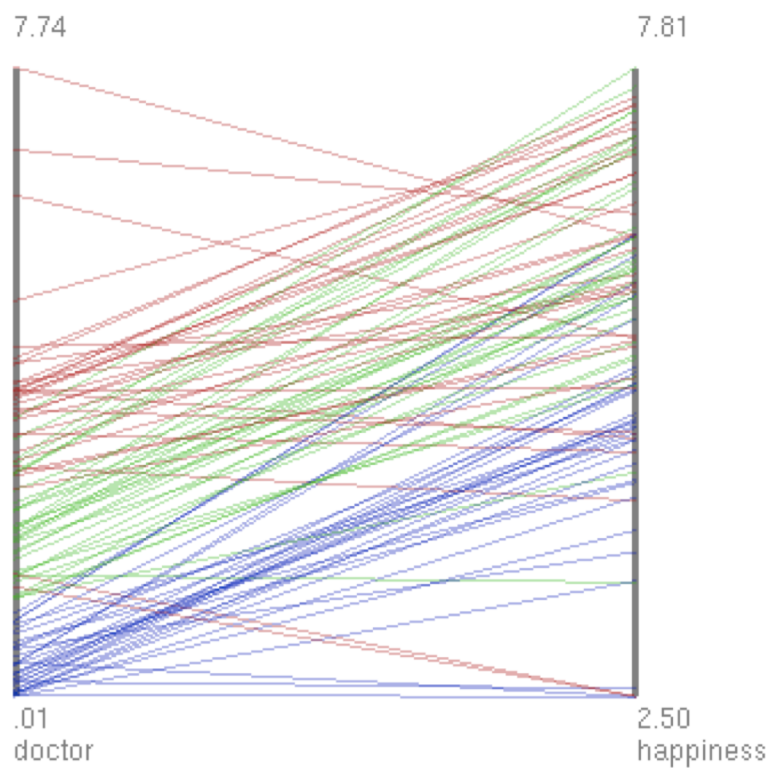
考察(喘息死亡者数と幸福度の関係)



負の相関がある

→前のグラフで幸福度が高位ほど先進国を指していて、また、喘息死亡者数が少ないほど先進国であることがわかっているのでこのグラフの形は予測できた。

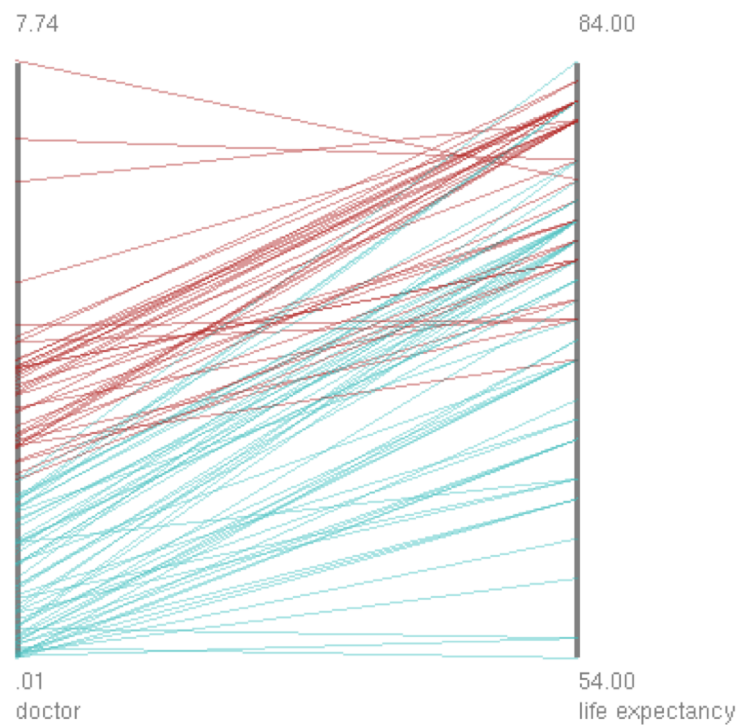
考察(幸福度と医師の数の関係)



正の相関がある

→医師の数が多いほど先進国を指していて、また、幸福度が高いほど先進国の傾向があることが前のグラフからわかっているなのでこの形も予測できる。

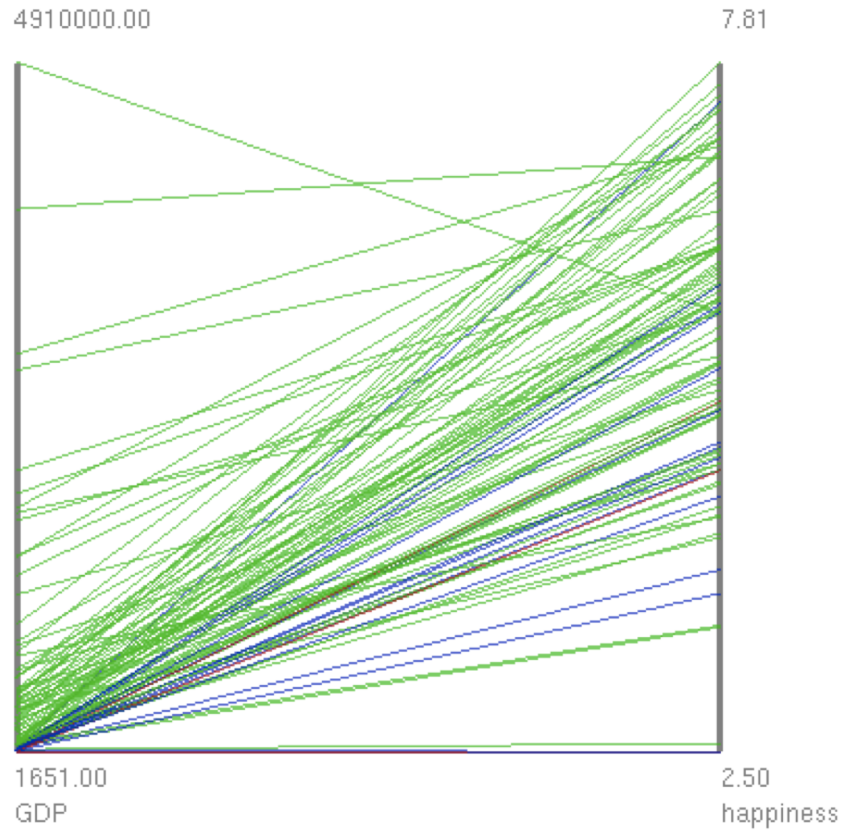
考察(医師の数と平均寿命の関係)



正の相関がある

→このグラフも医者が多いほど先進国で、平均寿命が高いほど先進国であることが前のスライドからわかっているので予測できる。

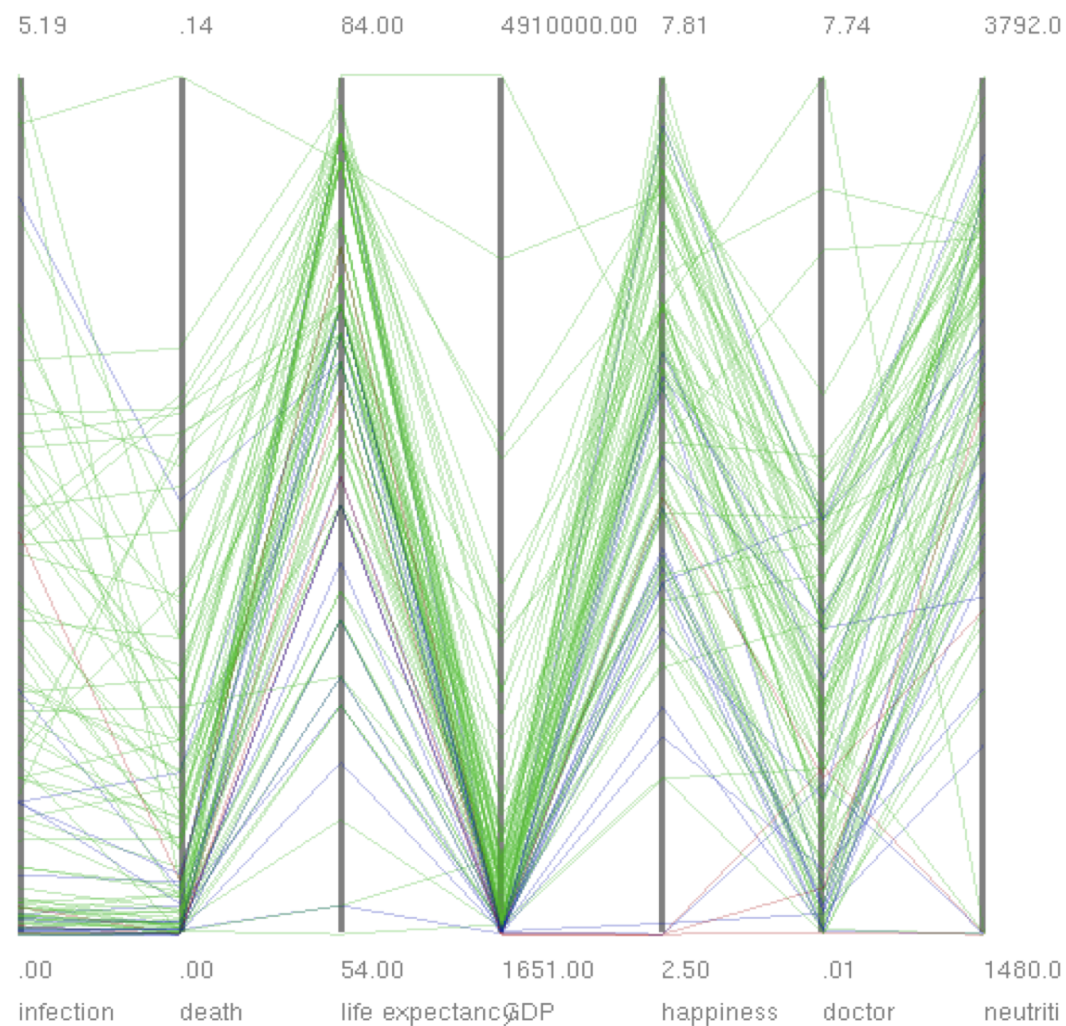
考察(GDPと幸福度の関係)



正の相関がある

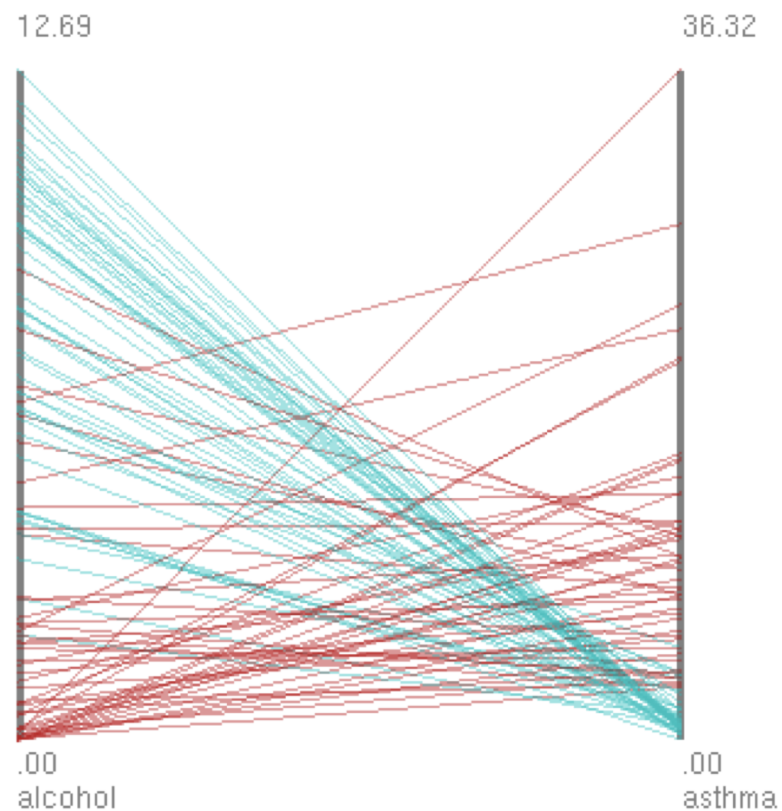
→GDPが高い国つまり先進国の方が幸福度が高いことは前のスライドからわかっているのでこのグラフも予測できる。

ここまでのまとめ



これらには強い正の相関関係があるので、散布図でも近い位置に全ての点が配置される。

考察(喘息死者数と飲酒量の関係)

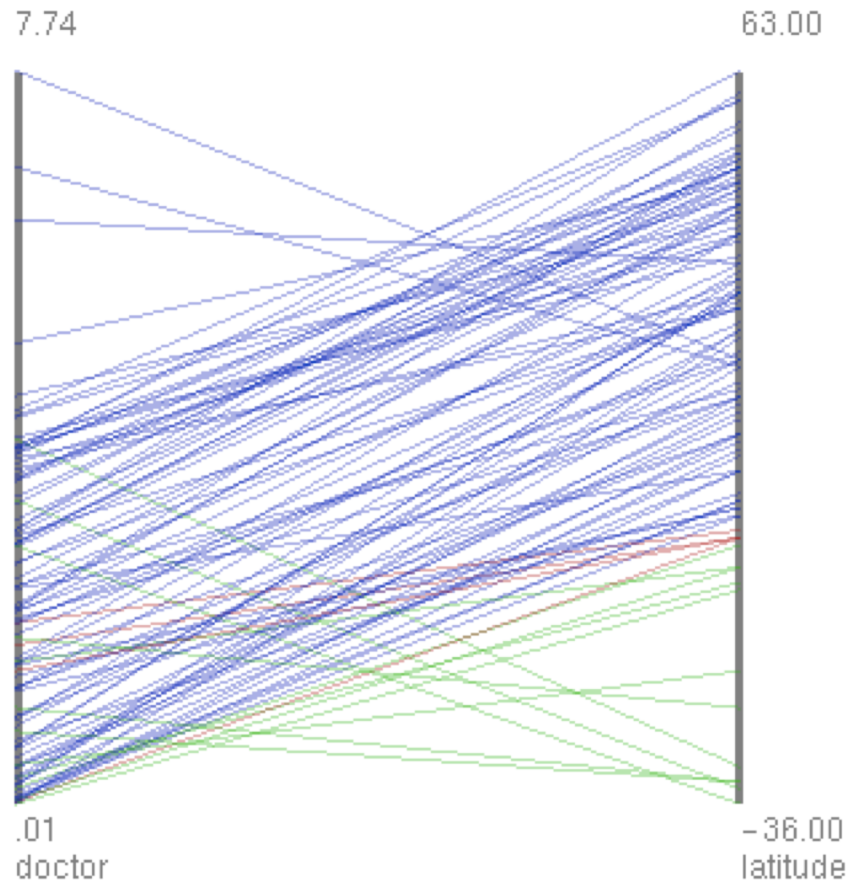


負の相関がある

→喘息死者数が少ないほど先進国であることがわかって
いるため、先進国の飲酒量が多いことがわかる。

飲酒にもお金がかかる為、裕福な家庭でないと沢山摂取する
ことは難しいことや、イスラム教徒は飲酒量が制限されている
ことが原因と考えられる。

考察(医師の数と緯度の関係)

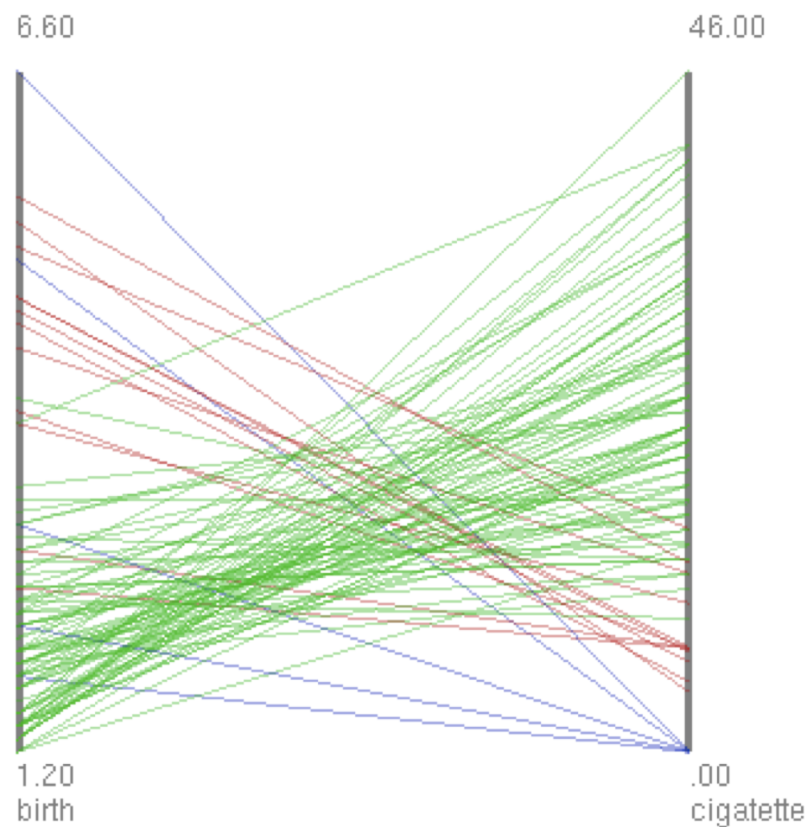


正の相関がある

→緯度が高いほど単位人数あたりの医師の数が多いことがわかる。

貧しい国が多いアジアやアフリカが赤道付近や南半球に位置することが原因と考えられる。

考察(出生数と喫煙率の関係)

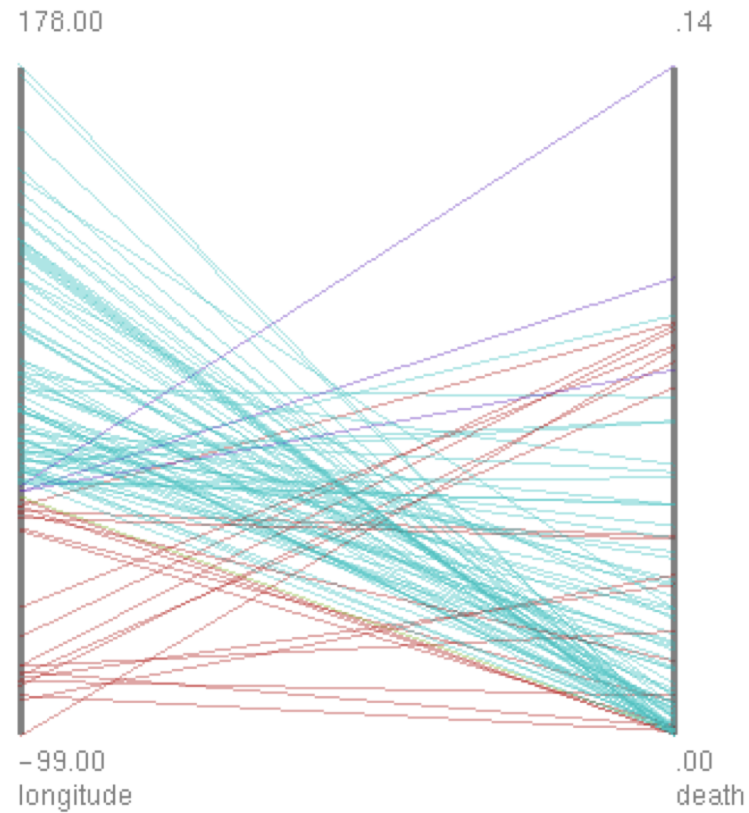


負の相関がある。

→出生数が多い、発展途上国ほど喫煙率が少ない。
特に、出生数が少ない先進国の喫煙率がとても高いことが読み取れる。

タバコを買う為の費用がかかり、生活費に影響を与えることが原因として考えられる。

考察(経度とコロナ死亡率の関係)

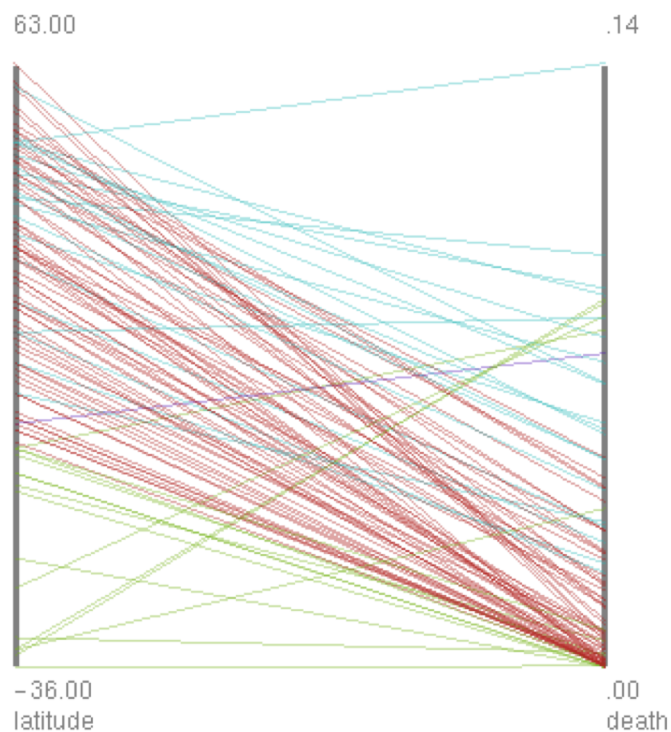


東経0~180度付近で負の相関がある

→東経が大きい地域ほど死亡者数が多いことがわかる。

原因は定かでない。

考察(緯度とコロナ死亡率の関係)

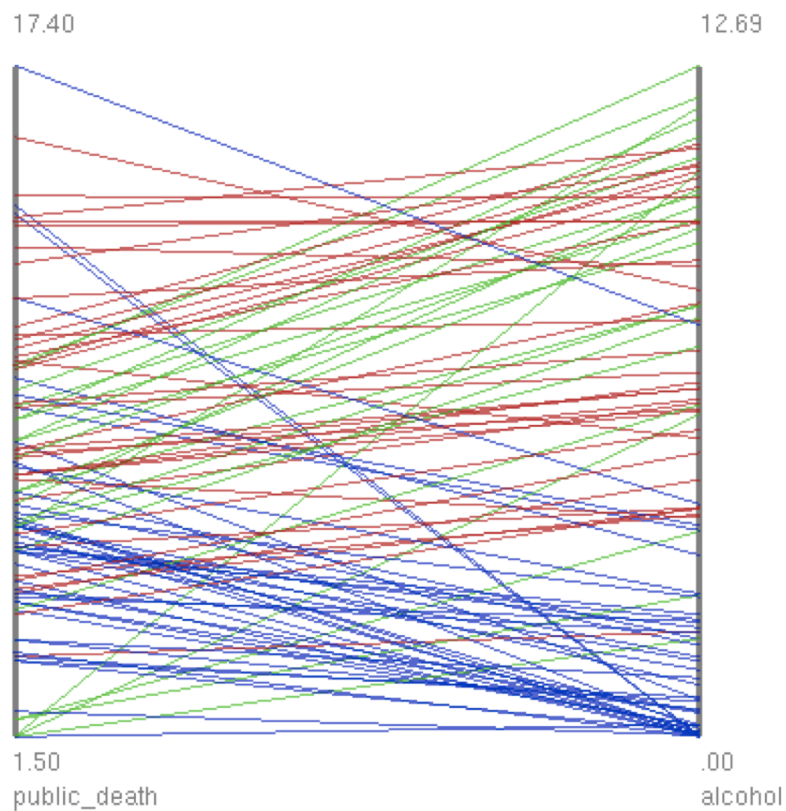


北緯0~90度付近で正の相関がある

→北半球で緯度が高い地域ほど死者数が多いことがわかる。

原因は定かでない。

考察(一般死亡率と飲酒量の関係)

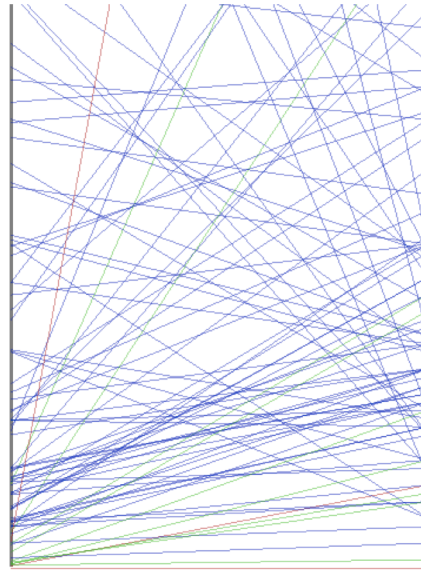
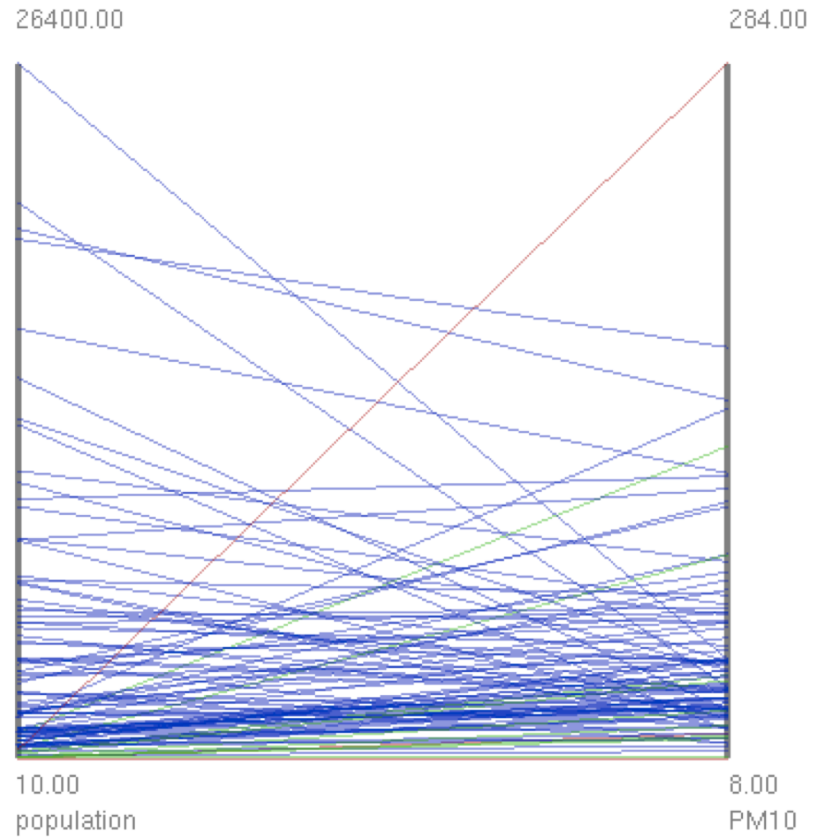


正の相関がある

→飲酒量が多い国ほど単位期間の1000人あたりの死亡者数が多い。

過度に飲酒をすると健康に悪影響を与えることが原因と考えられる。

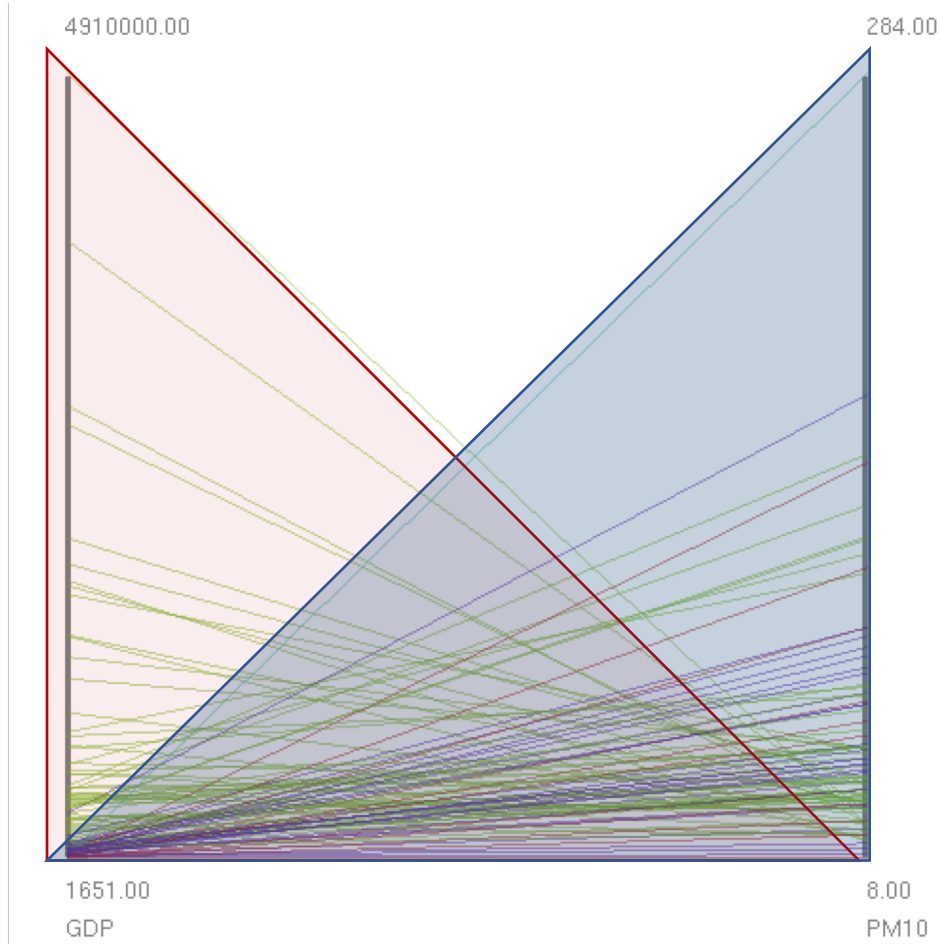
考察(PM10と人口の関係)



相関はほとんどない

→人口とPM10の量は関係がない。

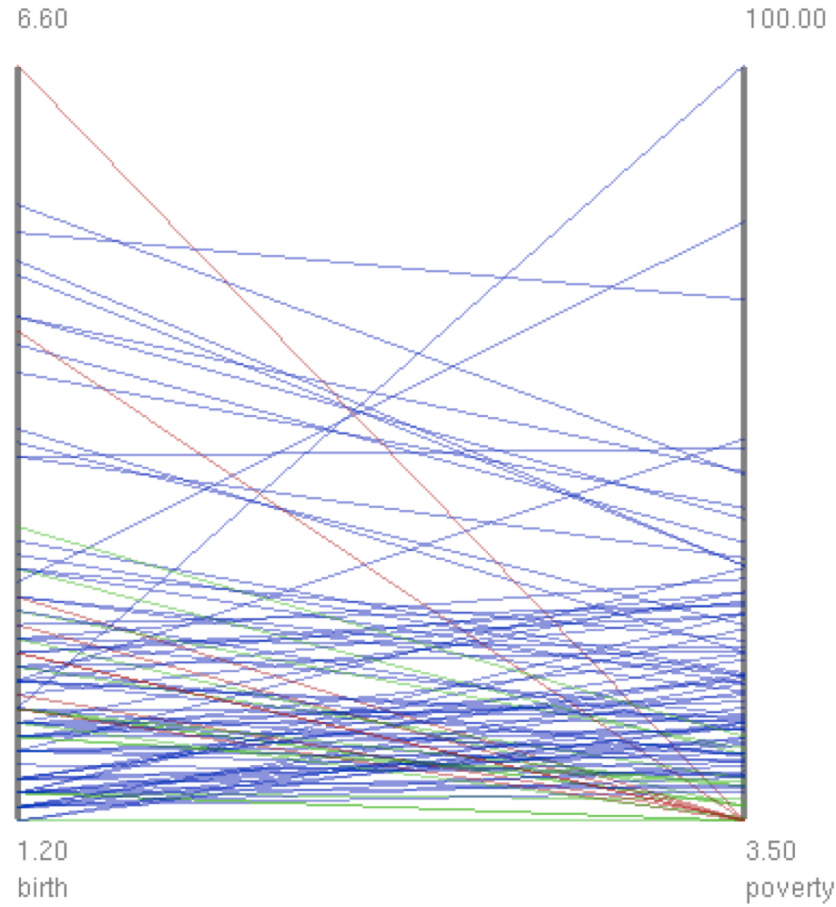
考察(PM10とGDPの関係)



2つの三角形の中でそれぞれ正の相関がある

→GDPが一定より大きい国、小さい国の中でそれぞれ正の相関がある。

考察(貧困率と出生数の関係)

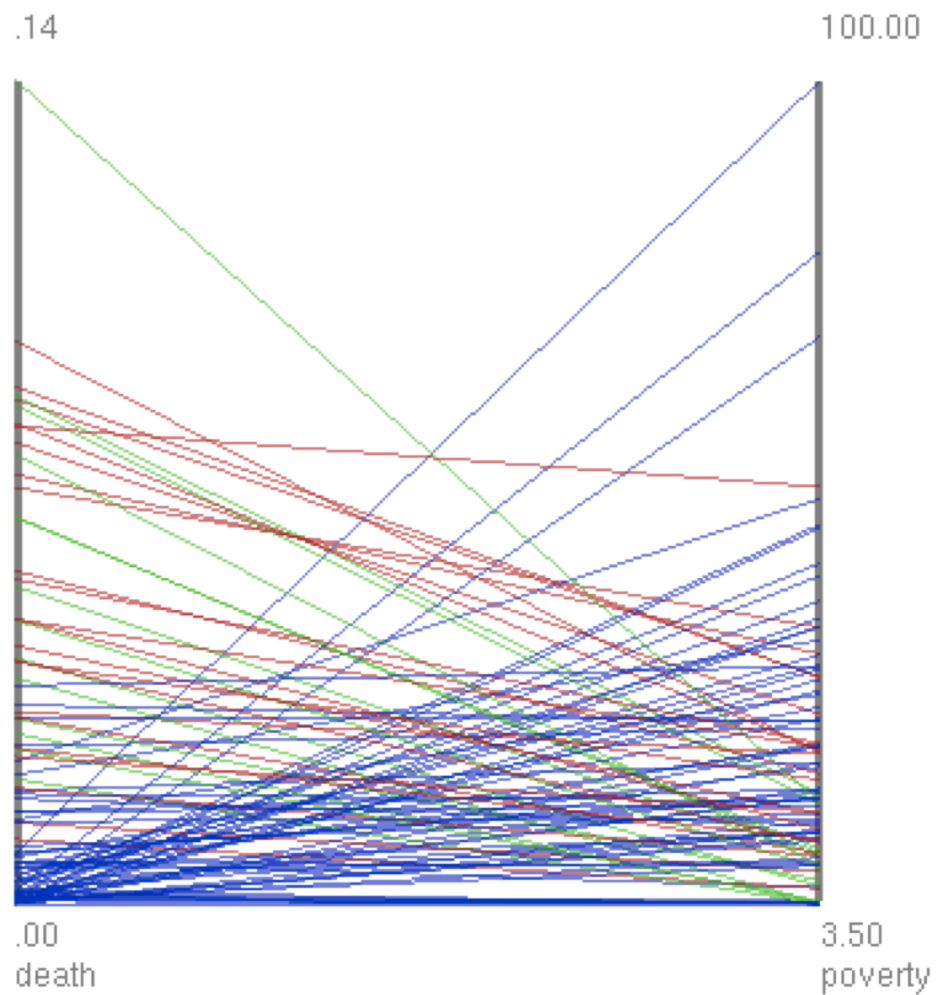


正の相関がある

→出生数が多い、つまり発展途上国ほど貧困率が高い傾向がある。

発展途上国ほど1日に使う金額が少ない人数が多いことが原因と考えられる。

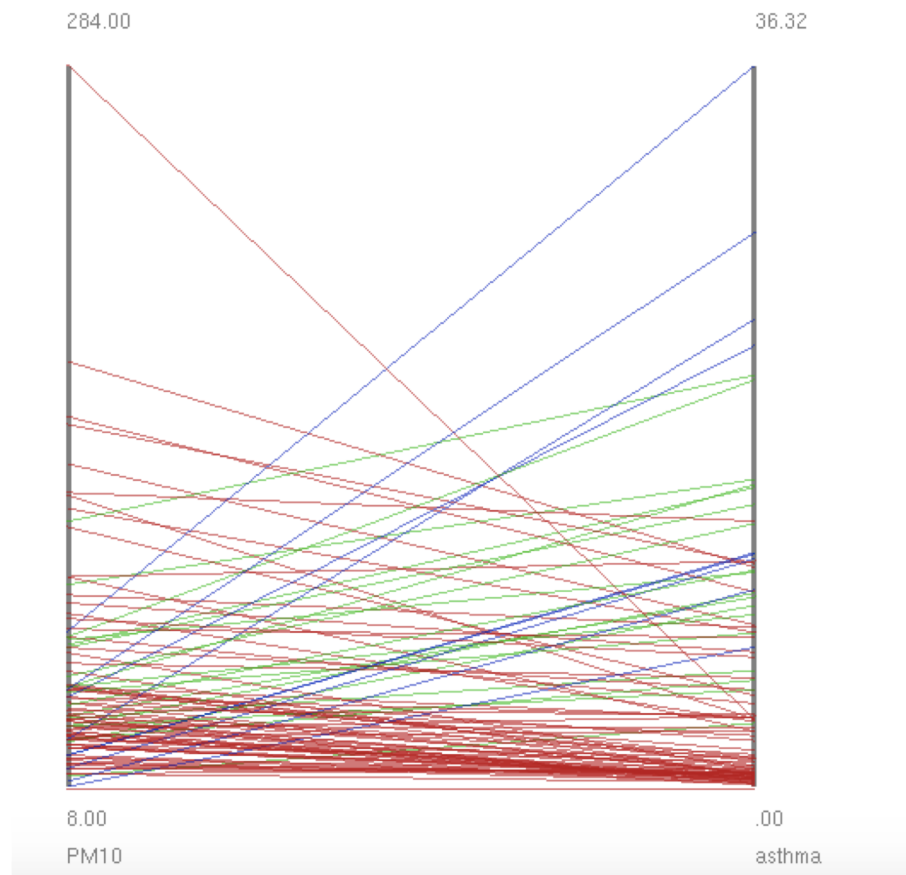
考察(貧困率とコロナ死亡率の関係)



負の相関がある

→コロナ死亡率が高い国つまり先進国ほど貧困率が低いことは前のスライドからわかっているなのでこの形は予測できる。

考察(PM10と喘息死亡者数の関係)

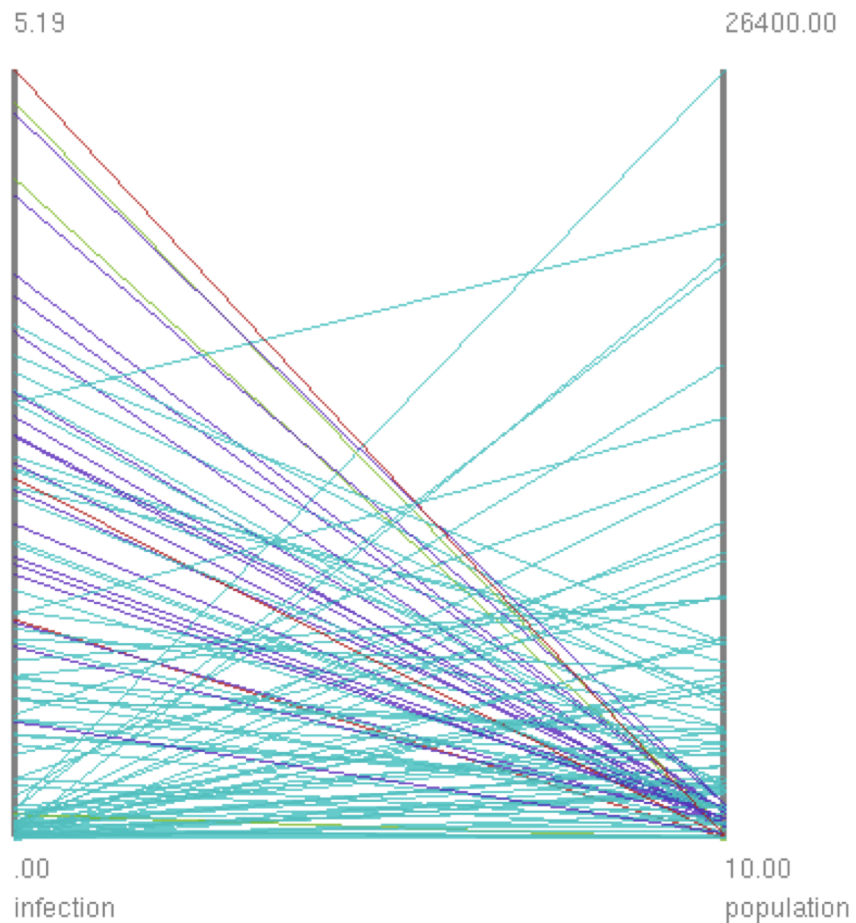


正の相関がある

→PM10が多い国ほど喘息死亡者数が多い傾向がある。

PM10は多くの異なる化合物から構成されていて、ブラックカーボンや油溶性がすを含む粒子性物質である為、体に悪影響を与えることが推測できる。

考察(コロナ感染率と人口の関係)



負の相関がある

→人口が多い国ほど感染率は低い。特に、グラフの紫の部分から、感染率が高い国の人口が少ないことがわかる。

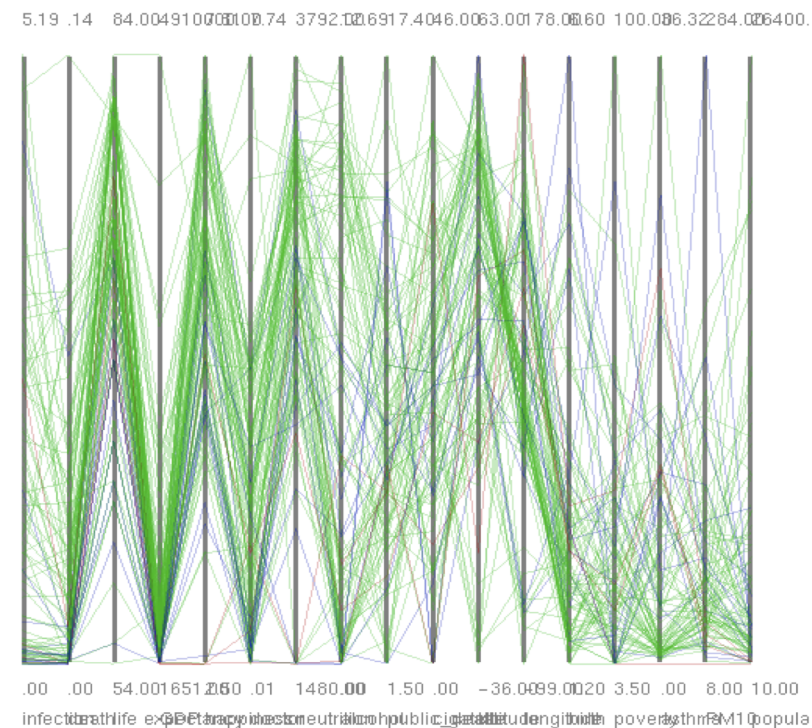
理由は定かでない。

まとめ

* コロナウイルス感染率、コロナウイルス死亡率と正の相関があるもの
GDP、平均寿命、幸福度、医師の人数、供給カロリー、飲酒率、一般死亡率、喫煙率
また、地球の北半球の東側では、北東部ほどコロナ死亡率が高い。

* コロナウイルス感染率、コロナウイルス死亡率と負の相関があるもの
出生率、貧困率、喘息死亡率

* どちらとも言えないもの
人口、PM10



(軸を入れ替えたもの)