

Asocjacja

- Nieprzypadkowe współwystępowanie czynników (alleli i fenotypów) na poziomie populacji
 - Czy zawsze asocjacja oznacza zależność przyczynową?
 - Czy każda asocjacja ma wartość diagnostyczną?
 - Czy asocjacja odkrywa “gen na”?

Asocjacje mogą być zwodne

- Allel 3A4 cytochromu P450 (CYP3A) i rak prostaty (bardziej zaawansowana postać w momencie diagnozy)
- CYP3A może w pływać na tempo hydroksylowania testosteronu - związek przyczynowy?
- Nie ma wpływu allelu 3A4 na kinetykę metabolizmu testosteronu
- Allel 3A4 częściej występuje u ludzi pochodzenia afrykańskiego (Afroamerykanie), niż europejskiego
- Podobne korelacje dla raka prostaty dla innych alleli częstszych w populacji afrykańskiej
- Bardziej zaawansowany rak prostaty u Afroamerykanów (przyczyny społeczne)?

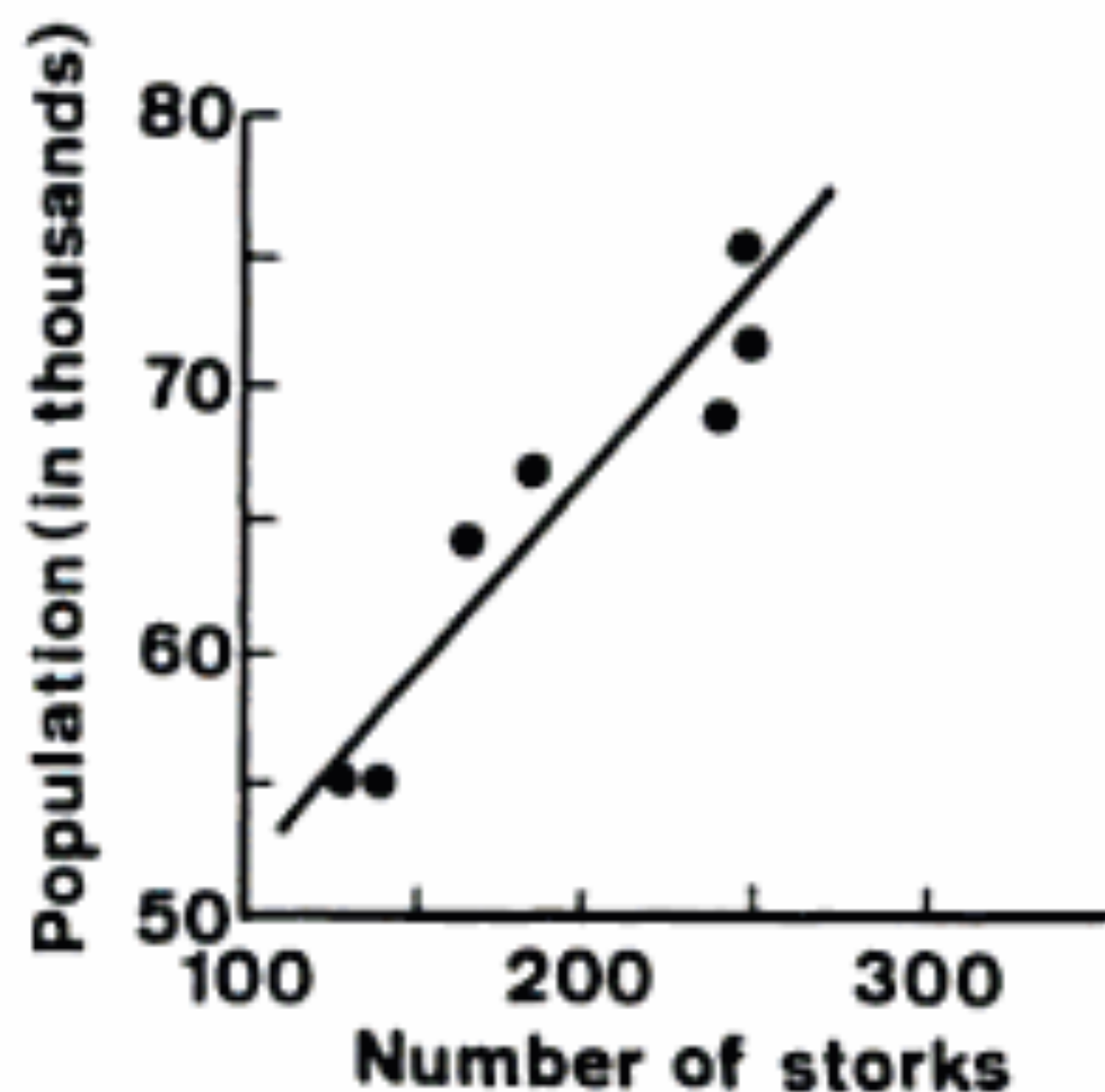
Ważne!!

- Asocjacja to nie jest “gen na...”!
- Czynniki ryzyka nie mają zwykle znaczenia diagnostycznego
 - Może być przydatny w diagnostyce różnicowej
- Zawsze należy analizować asocjację na tle ogólnego ryzyka w populacji, jakie są wartości bezwzględne

O korelacji i wynikaniu...

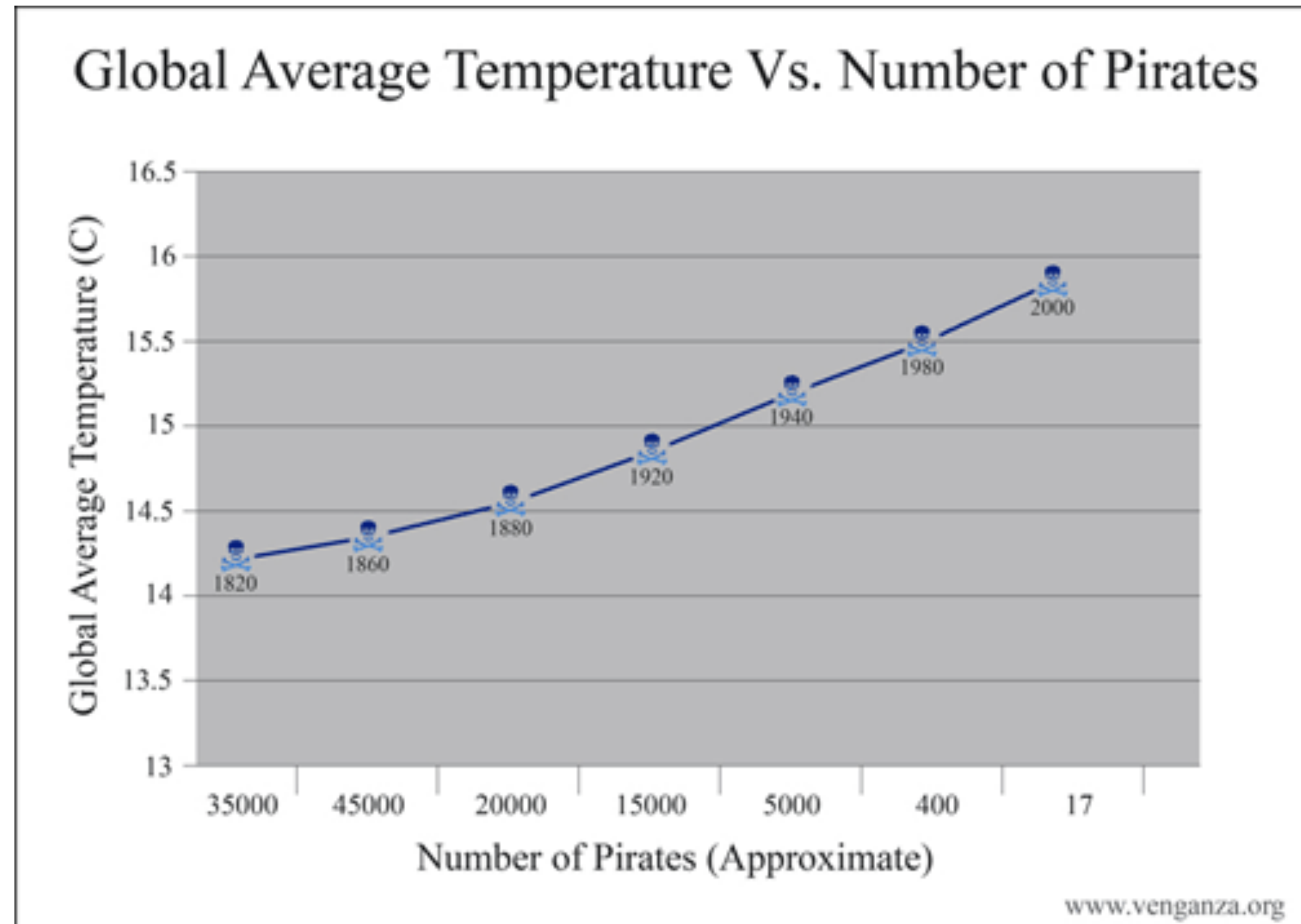
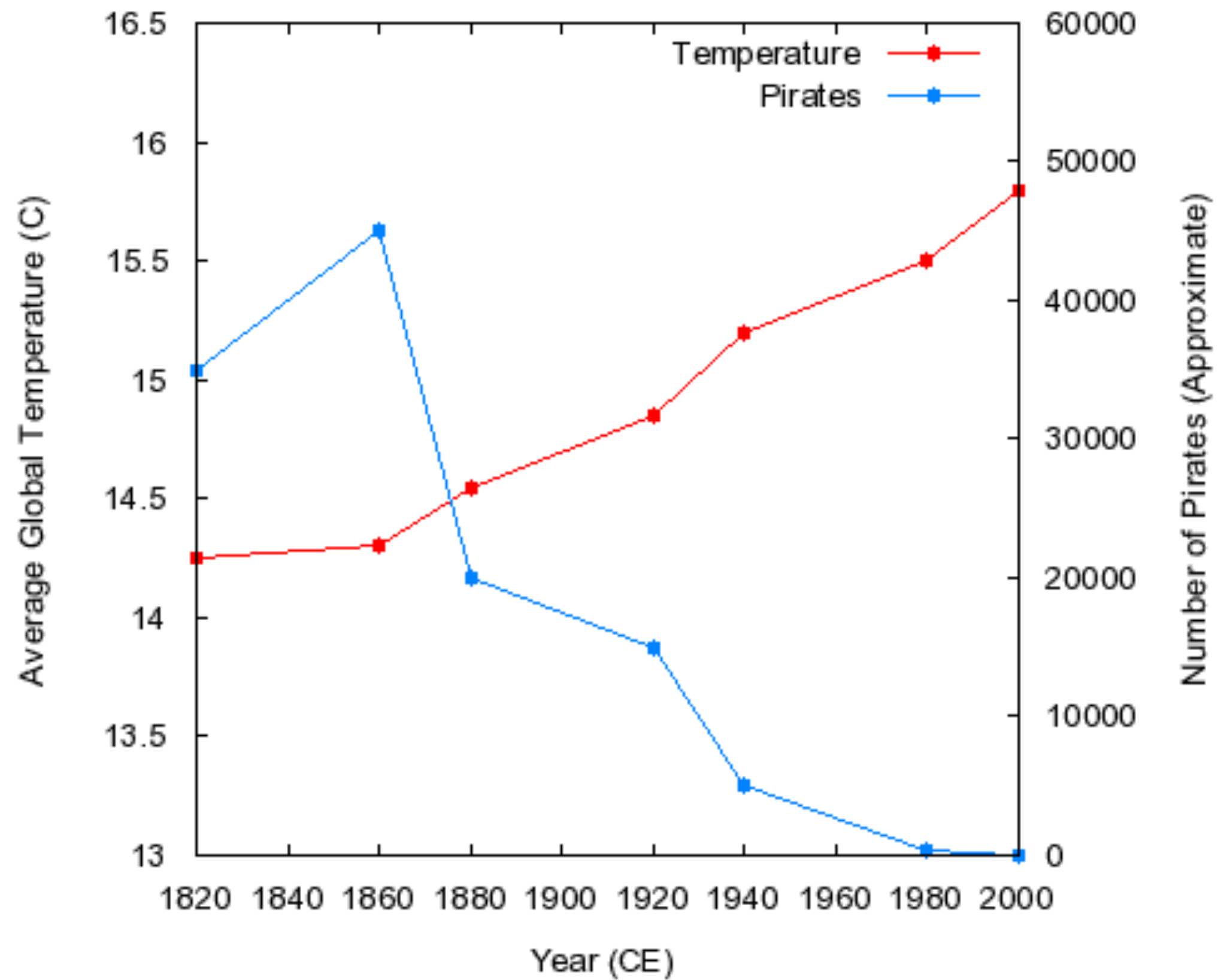


O korelacji i przyczynowości...

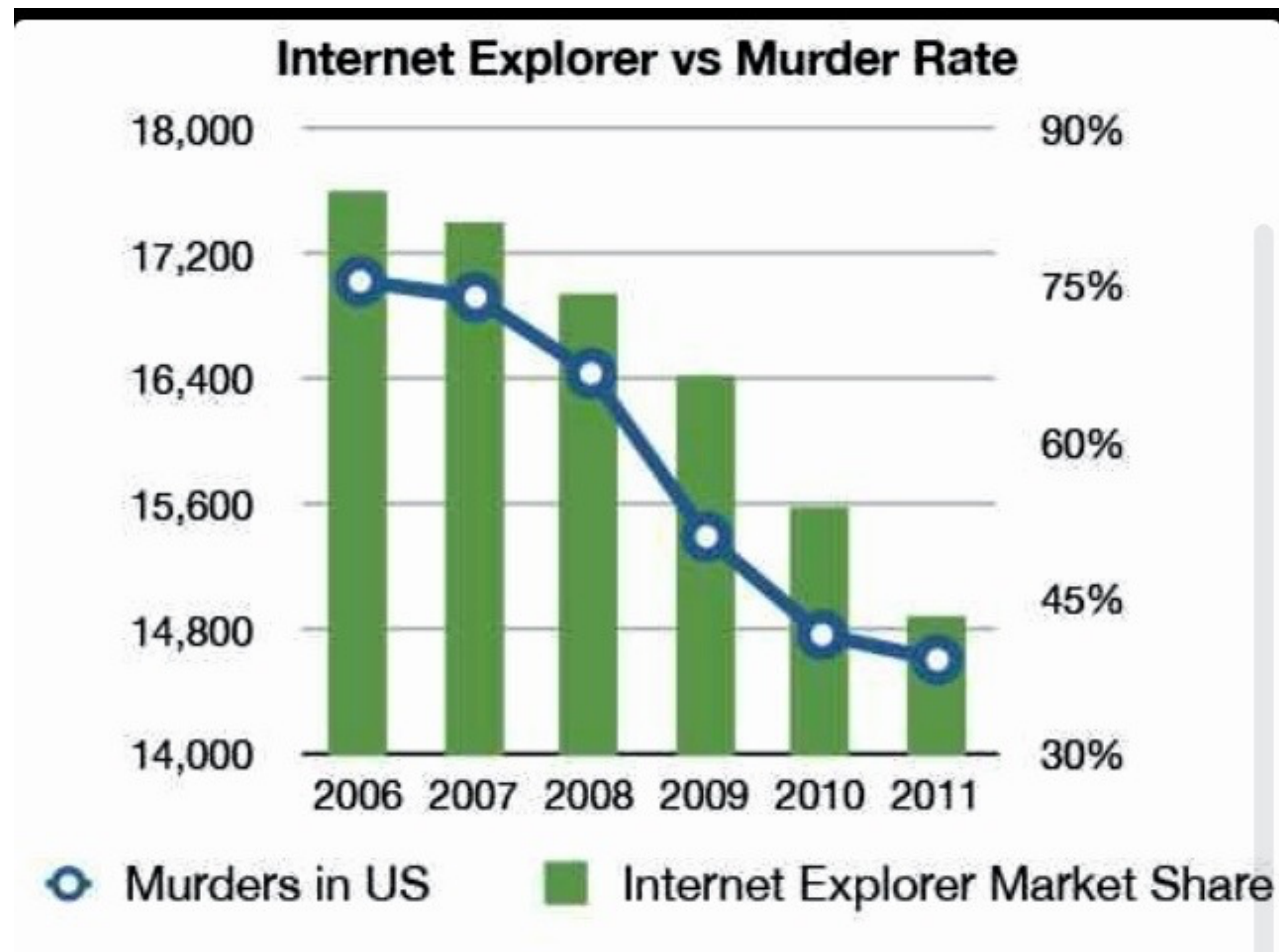


Populacja okolic Oldenburga, 1930-1936, za
Richard F. Mould "Introductory Medical Statistics"

O korelacji i przyczynowości...



O korelacji i przyczynowości...



Przyczyny asocjacji alleli

- Struktura populacji (ukryta zmienna)
- Paradoks Simpsona: Jeżeli zmieszamy dwie populacje, w których nie występuje asocjacja markera i choroby, ale są różne częstości występowania choroby (oraz markera), w zmieszanej populacji pojawi się asocjacja markera z chorobą (problem ukrytej zmiennej).
- Należy badać populacje jednorodnie etnicznie, a nawet tylko pojedyncze duże rodziny

Paradoks Simpsona - przykład

Prosty przykład: uniwersytet powinien faworyzować kobiety przy przyjmowaniu na studia doktoranckie:

	Mężczyźni	Kobiety
Historia	1/5 (20%)	2/8 (25%)
Geografia	6/8 (75%)	4/5 (80%)
Razem	7/13 (54%)	6/13 (46%)

Przykład medyczny

Porównywano dwie terapie kamieni nerkowych

Terapia A	Terapia B
78 % (273/350)	83% (289/350)

Ale:

	Terapia A	Terapia B
Małe kamienie	93% (81/87)	87% (234/270)
Duże kamienie	73% (192/263)	69% (55/80)
Razem	78 % (273/350)	83% (289/350)

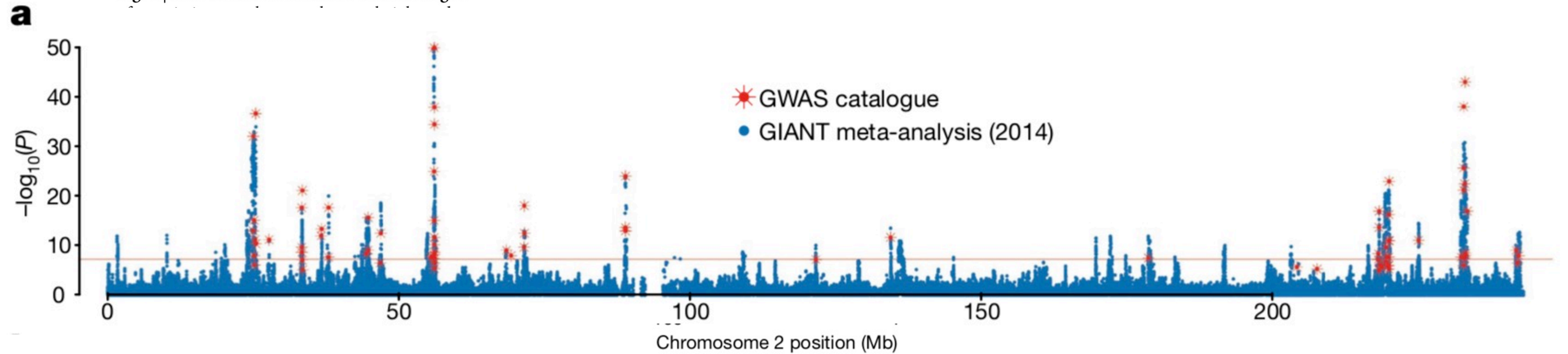
Terapię A częściej stosowano u pacjentów z dużymi kamieniami, które trudniej się leczy.

Genetyczne czynniki ryzyka

- Badania GWAS (*Genome-Wide Association Studies*) – asocjacje na skalę genomu
- Korelacja wariantów genetycznych z ryzykiem choroby
- Znajduje się wiele korelacji, ale żadna nie może być decydującym czynnikiem
- Tajemnica brakującej odziedziczalności

Wizualizacja GWAS - Manhattan plot

Fig. 4 | Association statistics for human height.



The UK Biobank resource with deep phenotyping and genomic data

Clare Bycroft^{1,13}, Colin Freeman^{1,13}, Desislava Petkova^{1,12,13}, Gavin Band¹, Lloyd T. Elliott², Kevin Sharp², Allan Motyer³, Damjan Vukcevic^{3,4}, Olivier Delaneau^{5,6,7}, Jared O'Connell⁸, Adrian Cortes^{1,9}, Samantha Welsh¹⁰, Alan Young¹¹, Mark Effingham¹⁰, Gil McVean^{1,11}, Stephen Leslie^{3,4}, Naomi Allen¹¹, Peter Donnelly^{1,2,14} & Jonathan Marchini^{1,2,14*}

Istotność wartości p w GWAS

- $p < 5 \cdot 10^{-8}$ próg zwyczajowy (dla populacji europejskiej)
- można też wyliczyć istotność korzystając z różnych poprawek statystycznych
 - Bonferroniego (prosta, ale zbyt restrykcyjna)
 - Benjaminiego–Hochberga

Analiza asocjacji

Dane:

- rodowód i genotypy: plik .ped
- mapa: plik .map

Dane w formacie binarnym

- rodowód: plik .fam (pierwsze 6 kolumn pliku .ped)
- mapa: plik .bim (dodatkowe dane: allele i MAF)
- genotypy markera: plik .bed (skompresowana wersja pliku tekstowego)

Inne częste formaty

VCF

```
##fileformat=VCFv4.2
##contig=<ID=2,length=51304566>
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE1 SAMPLE2 SAMPLE3 SAMPLE4 SAMPLE5 SAMPLE6 SAMPLE7
2 81170 . C T . . AC=9;AN=7424 GT:DP:GQ 0/0:4:12 0/0:3:9 0/1:1:3 0/1:9:24 1/0:4:12 0/0:5:15 0/0:4:12
2 81171 . G A . . AC=6;AN=7446 GT:DP:GQ 0/1:4:12 0/0:3:9 0/0:1:3 0/0:9:24 0/1:4:12 0/1:5:15 0/0:4:12
2 81182 . A G . . AC=5;AN=7506 GT:DP:GQ 0/0:5:15 0/0:4:12 0/0:5:15 0/0:9:24 0/0:4:12 0/0:4:12 0/0:4:12
2 81204 . T G . . AC=2;AN=7542 GT:DP:GQ 1/0:5:15 0/0:9:27 0/0:10:30 0/0:15:39 0/0:9:27 1/0:13:39 0/1:14:42
```

BCF

```
2 81170 . C T . . AC=9;AN=7424 GT:0/0:0/0:0/1:0/1:1/0:0/0:0/0 DP:4:3:1:9:4:5:4 GQ:12: 9: 3:24:12:15:12
2 81171 . G A . . AC=6;AN=7446 GT:0/1:0/0:0/0:0/0:0/1:0/1:0/0 DP:4:3:1:9:4:5:4 GQ:12: 9: 3:24:12:15:12
2 81182 . A G . . AC=5;AN=7506 GT:0/0:0/0:0/0:0/0:0/0:0/0:0/0 DP:5:4:5:9:4:4:4 GQ:15:12:15:24:12:12:12
2 81204 . T G . . AC=2;AN=7542 GT:1/0:0/0:0/0:0/0:0/0:1/0:0/1 DP:5:9:10:15:9:13:14 GQ:15:27:30:39:27:39:42
```



plink (w wersji 1.9b lub wyższej) odczytuje pliki vcf i bcf i przekształca na .bed/.bim/.fam

Program plink

`plink --file nazwa`

wczytuje pliki nazwa.ped nazwa.map

`plink --bfile nazwa`

wczytuje pliki nazwa.fam nazwa.bim nazwa.bed

Konwersja formatów

```
plink --file przyklad --make-bed --out nowanazwa
```

```
plink --bfile przyklad --recode --tab --out nowanazwa
```

Przykładowe możliwości plink

`--assoc`

analiza asocjacji

`--assoc --adjust`

j.w. z poprawkami statystycznymi

`--out nowanazwa`

Przykładowe możliwości plink

`--keep / --remove lista`

zachowuje / usuwa osoby o ID z pliku lista

`--keep-fam --remove-fam`

j.w., tylko dla ID rodziny a nie osoby

`--extract / --exclude lista`

zachowuje / usuwa markery o nazwach z pliku lista

+ wiele innych (filtrowanie wg. chromosomów, MAF, SNP vs. indele, i wiele innych)

`--make-bed --out nowanazwa`

zapisuje efekt filtrowania w nowym pliku

Interpretacja wg. bazy genomu człowieka

- Na stronie ncbi.nlm.nih.gov, w sekcji *Genetics & Medicine*, baza ClinVar, narzędzie *Variation Viewer*
- Formaty zapytania (przykłady)
 - rs28693734 - identyfikator SNP (z bazy dbSNP), o ile dostępny
 - chr1:215210111 - chromosom:pozycja (uwaga na genom referencyjny)

Analiza z projektu 1000 genomów

```
study.bed study.bim study.fam
```

```
--allow-no-sex
```

konieczne aby uwzględnione zostały osoby bez informacji o płci

- Wykres manhattan do pliku jpg (otwieramy komendą `jpeg("nazwa")`), zamykamy `dev.off`