# Przetwarzanie danych środowiskowych

## Analiza korelacji

# 1. Analiza związku dwóch zmiennych: korelacja

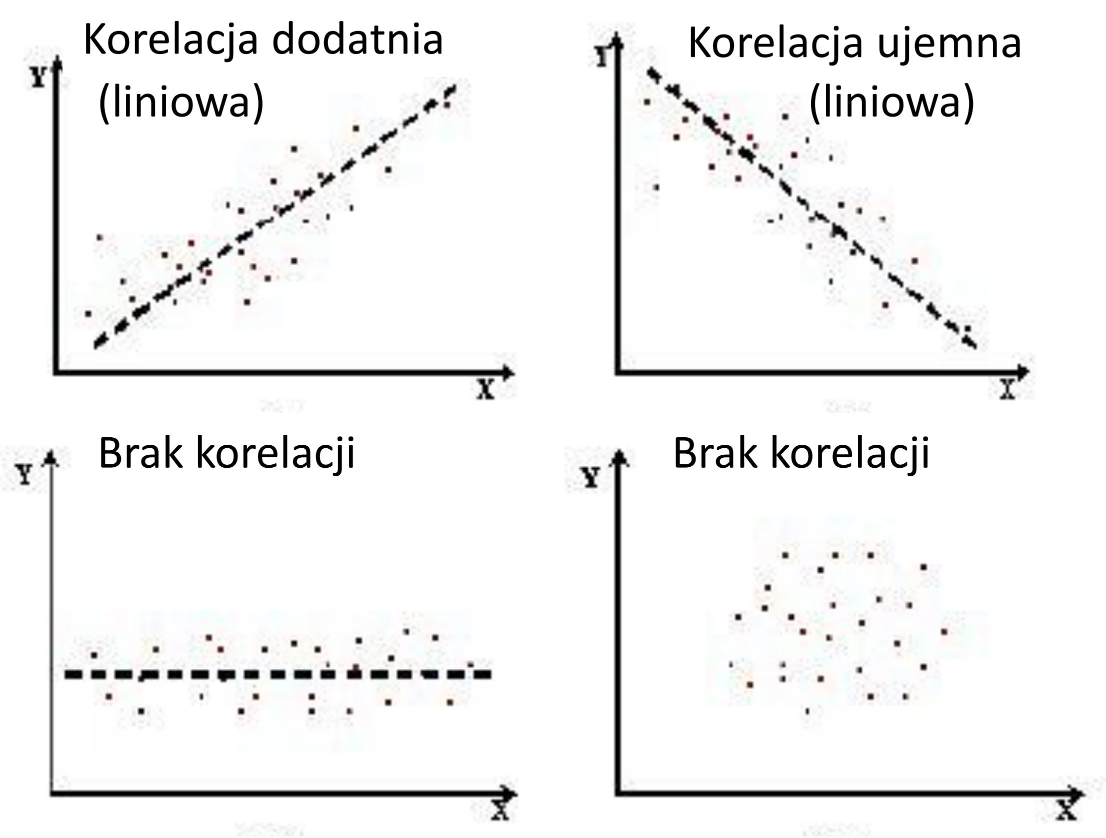
**Korelacja** (współzależność zmiennych) określa **wzajemne powiązania pomiędzy** wybranymi **zmiennymi**.
Charakteryzując korelację dwóch zmiennych podajemy dwa czynniki: **kierunek** oraz **siłę**.

Cel analizy korelacji ->

stwierdzenie, czy między badanymi

zmiennymi :

- zachodzą jakieś zależności,
- jaka jest siła zależności,
- jaka jest postać zależności,
- jaki jest kierunek zależności.



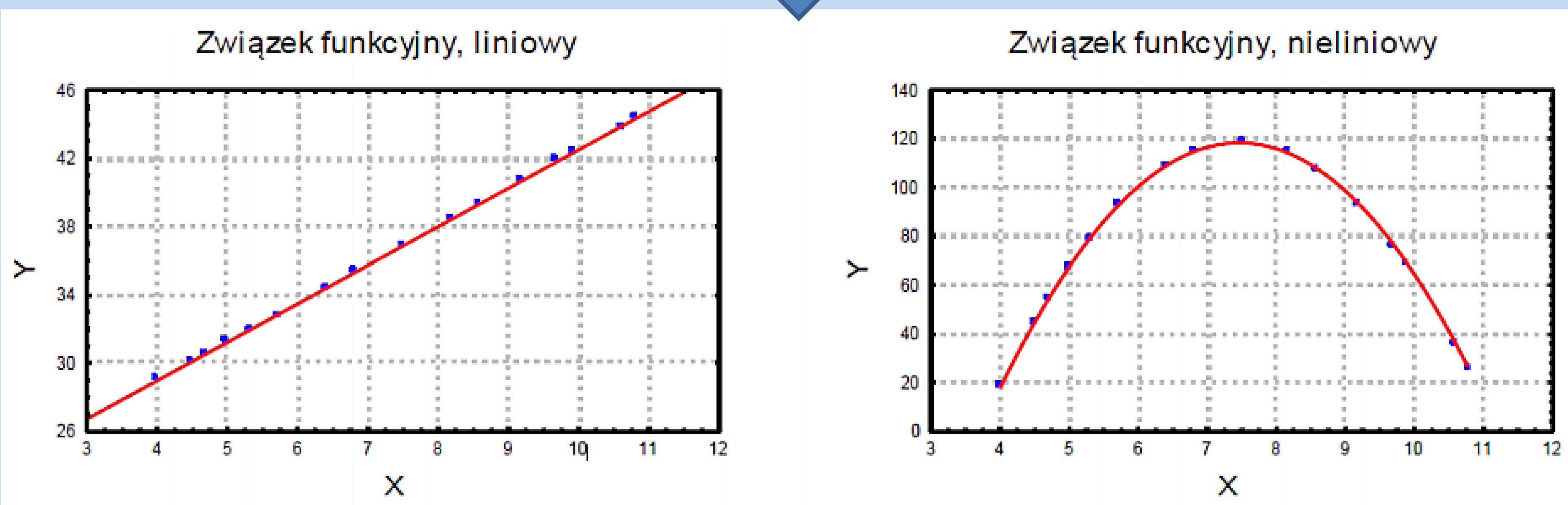

Khelil i in. 2007

# 1. Analiza związku dwóch zmiennych: korelacja

Współzależność między zmiennymi może być dwojakiego rodzaju:

- **funkcyjna** -> zmiana wartości jednej zmiennej powoduje ściśle określoną zmianę wartości drugiej zmiennej.

W przypadku zależności funkcyjnej, określonej wartości jednej zmiennej (X) odpowiada jedna i tylko jedna wartość drugiej zmiennej (Y).





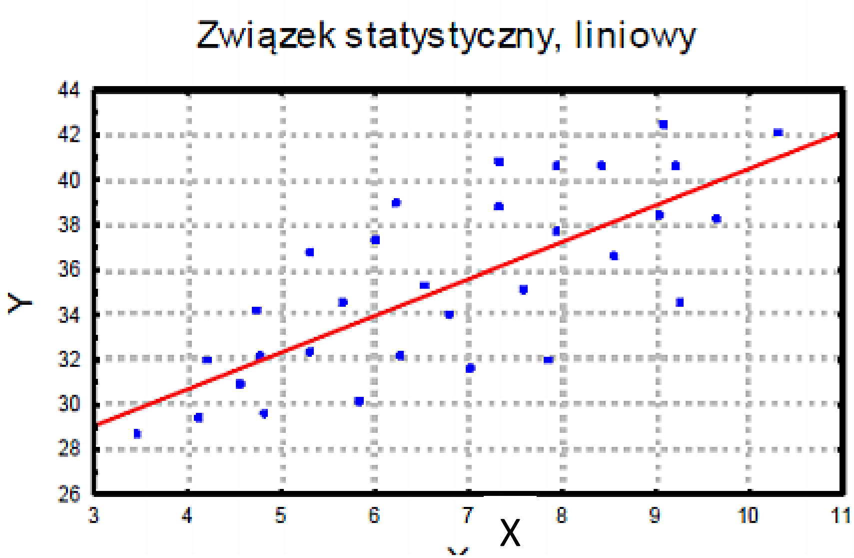

http://zsi.tech.us.edu.pl/~nowak/odzw/SMAD_korelacje.pdf

# 1. Analiza związku dwóch zmiennych: korelacja

- **stochastyczna (probabilistyczna)** -> wraz ze zmianą wartości jednej zmiennej zmienia się rozkład prawdopodobieństwa drugiej zmiennej.

  Szczególny przypadek zależności stochastycznej -> określonym wartościom jednej zmiennej odpowiadają ściśle określone <u>średnie wartości</u> drugiej zmiennej.

  Możemy wtedy ustalić, jak zmieni się – średnio biorąc – wartość zmiennej zależnej Y (inaczej objaśnianej) w zależności od wartości zmiennej niezależnej X (inaczej objaśniającej)*.





http://zsi.tech.us.edu.pl/~nowak/odzw/SMAD_korelacje.pdf

*Przewidywanie wartości zmiennych objaśnianych (*y*) na podstawie wartości zmiennych objaśniających (*x*) jest możliwe dzięki znalezieniu tzw. **modelu regresji**. W praktyce, w przypadku regresji liniowej, polega to na podaniu równania prostej, zwanej prostą regresji.



zmienne niezależne i zależne – zobacz np. tutaj: https://www.naukowiec.org/wiedza/metodologia/zmienne-niezalezne-i-zalezne_652.html

# 3. Współczynniki korelacji.

Miarą związku (wyrazem liczbowym) między zmiennymi są **współczynniki korelacji.**

- współczynniki korelacji przyjmują wartości z przedziału **[-1, +1],**
- wartość **-1** reprezentuje doskonałą korelację ujemną,
- wartość **+1** reprezentuje doskonałą korelację dodatnią,
- wartość **0** to brak korelacji.

Do analizy zależności ilościowych dwóch zmiennych najczęściej stosuje się:

- **współczynniki korelacji liniowej (np. Pearsona)**

(zależność liniowa – szczególny przypadek zależności monotonicznej, w której jedna ze zmiennych jest liniowo zależna od drugiej zmiennej)

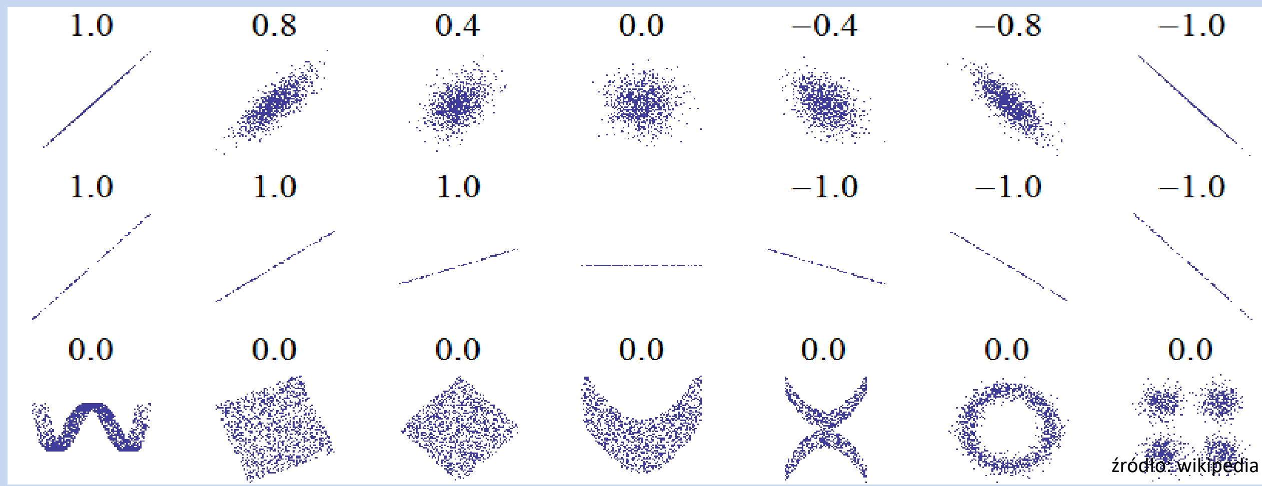
- **współczynniki korelacji monotonicznej (np. Spearmana)**

(zależność monotoniczna – gdy zwiększenie wartości jednej ze zmiennych oznacza zwiększenie lub zmniejszenie wartości oczekiwanej drugiej zmiennej)

# 2. Korelacyjne wykresy rozrzutu.

Zanim obliczymy wartość współczynników korelacji pomiędzy dwoma zmiennymi warto narysować ich wykres rozrzutu, pozwoli to m.in. uniknąć błędów związanych z wyborem odpowiedniej metody korelacji oraz interpretacji uzyskanych wyników korelacji.

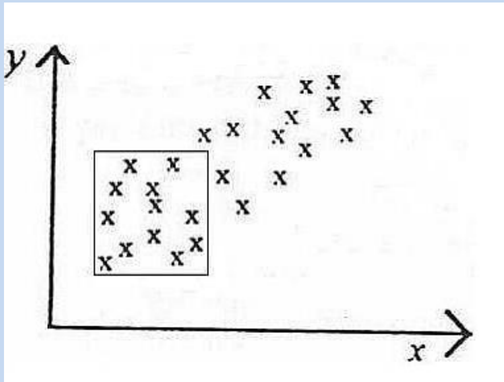
Wykresy rozrzutu prezentujące przykładowe zależności pomiędzy dwoma zmiennymi wraz z odpowiadającymi im wartościami współczynnika korelacji Pearsona



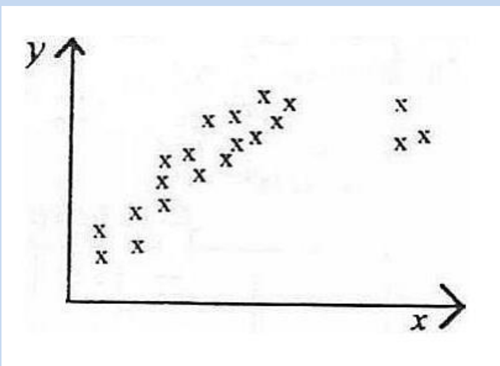

*Zauważ, że dla wszystkich wykresów z ostatniej linii wartość współczynnika korelacji Pearsona wynosi 0 . Nie oznacza to jednak, że nie ma związku między zmiennymi (wykresy wskazują istnienie jakiegoś związku), ale że korelacja ta nie jest liniowa (współczynnik korelacji Pearsona to współczynnik korelacji liniowej!!!). Dlatego przy analizie korelacji zawsze konieczna jest analiza korelogramu (czyli nie tylko opieranie się na wartości współczynnika korelacji).

# 2. Korelacyjne wykresy rozrzutu.

- błędy we wnioskowaniu o zależności pomiędzy X i Y

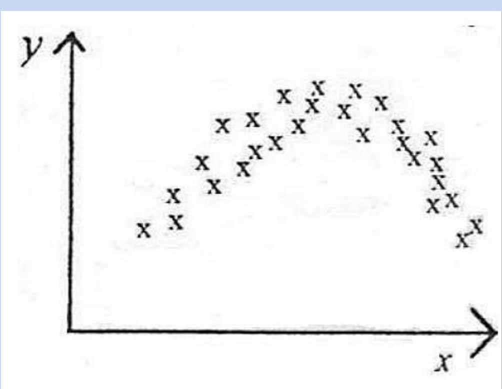



Za mało danych. Gdybyśmy narysowali korelogram tylko na podstawie niewielkiej liczby danych (reprezentowanych przez wartości obwiedzione kwadratem), wnioskowalibyśmy, że korelacji brak. W przypadku większej liczby danych okazuje się, że zależność jest liniowa.




Nietypowe dane. Trzy odseparowane punkty to dane nietypowe. Sugerują zależność nieliniową miedzy zmiennymi X i Y (parabola). Po odrzuceniu tych nietypowych wartości widać wyraźną zależność liniową.

- zależność nieliniowa




Smuga punktów układa się w kształt paraboli. Istnieje zależność pomiędzy zmiennymi X i Y i jest to związek nieliniowy (zależność nieliniowa).



wykorzystano: http://zsi.tech.us.edu.pl/~nowak/odzw/SMAD_korelacje.pdf

# 2. Korelacyjne wykresy rozrzutu.

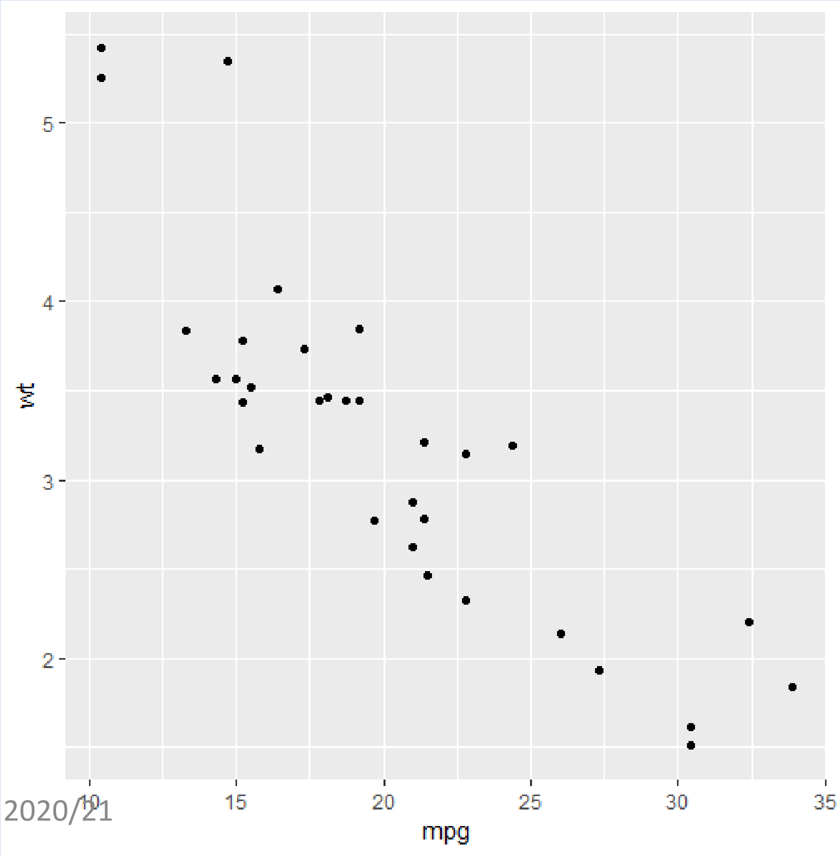
Kwartet Anscombe’a - dlaczego wykonując analizę danych warto narysować wykresy rozrzutu:
https://pl.wikipedia.org/wiki/Kwartet_Anscombe%E2%80%99a

# 2. Korelacyjne wykresy rozrzutu.

Narysujmy wykres rozrzutu dla zmiennych `mpg` i `wt` (zbiór danych `mtcars`):

```
> ggplot(mtcars, aes(mpg, wt))+ geom_point()
```

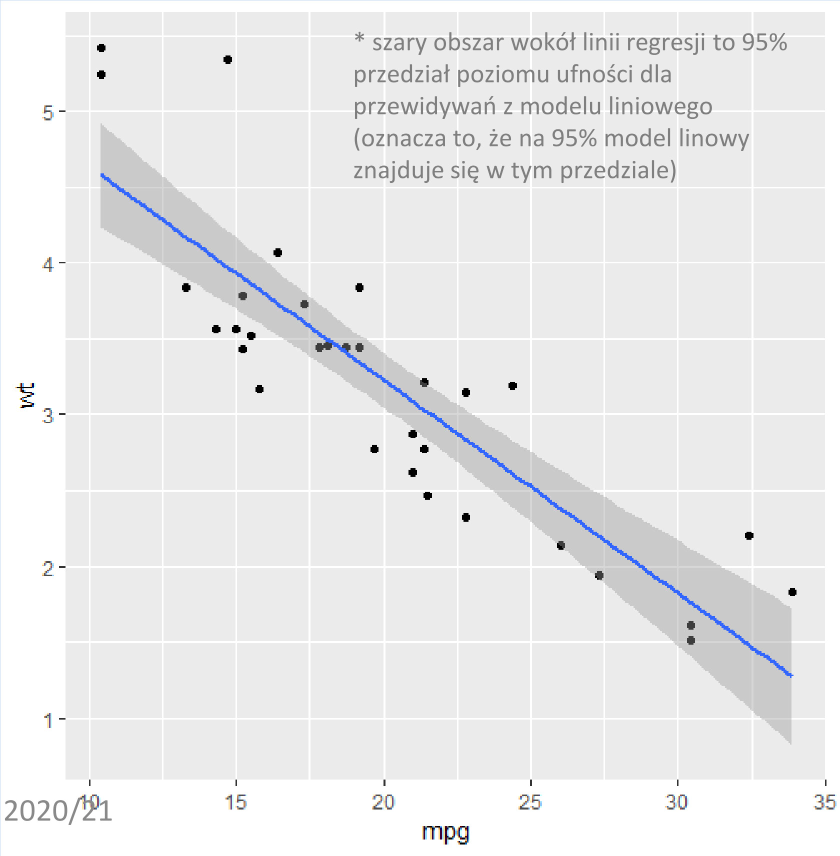
Możemy dopatrzeć się pewnej zależności liniowej w danych. Czy jest to zależność dodatnia czy ujemna?



# 2. Korelacyjne wykresy rozrzutu.

Dodajmy do wykresu linię trendu (prostą regresji):

```
> ggplot(mtcars, aes(mpg, wt))+ geom_point()+ geom_smooth(method="lm")
```

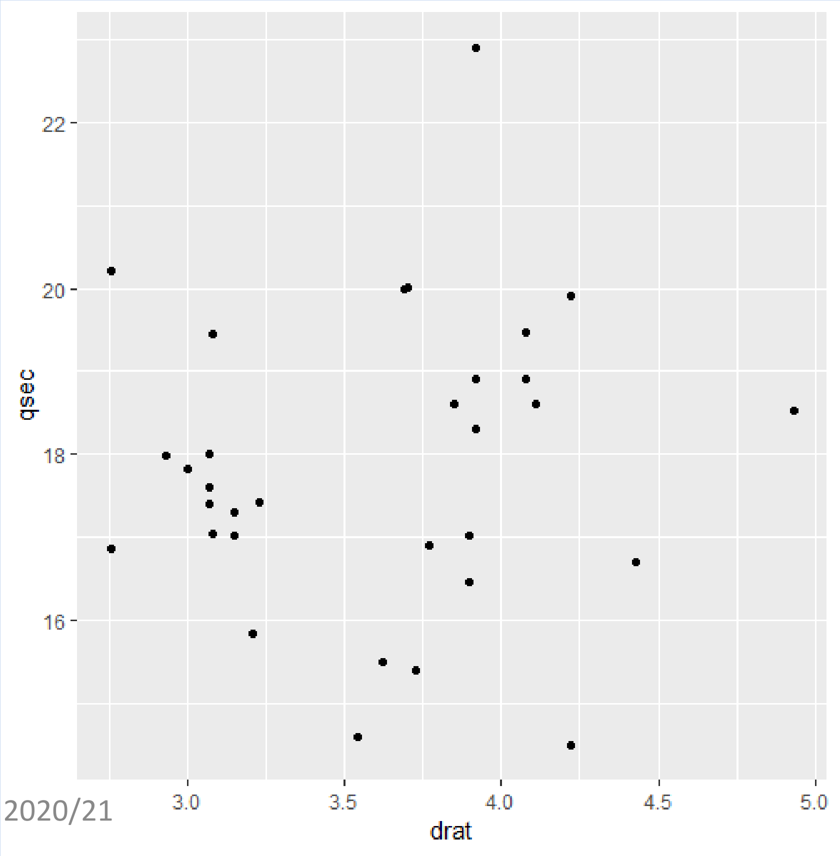


# 2. Korelacyjne wykresy rozrzutu.

Zadanie: Narysuj wykres rozrzutu dla zmiennych `drat` i `qsec` (zbiór danych `mtcars`):

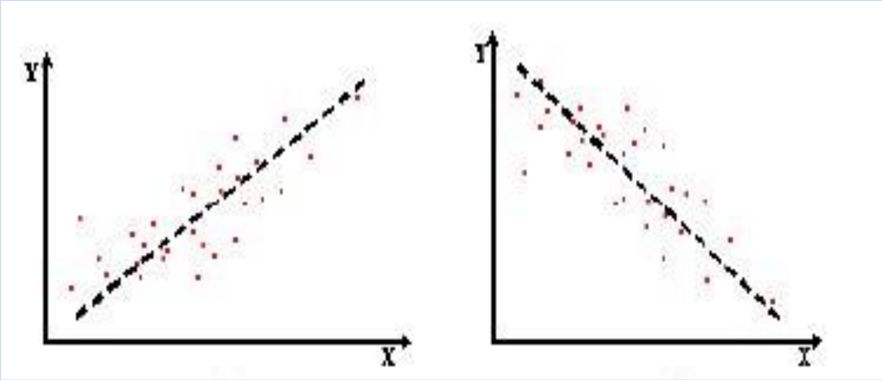
```
> ggplot(mtcars, aes(drat, qsec))+ geom_point()
```

Czy możemy dopatrzeć się tutaj zależności liniowej?



# 3.1. Współczynnik korelacji Pearsona.

Wspomniany już wcześniej **współczynnik korelacji Pearsona (*r*)** wykorzystywany jest do badania związków prostoliniowych badanych zmiennych, w których zwiększenie wartości jednej z cech powoduje proporcjonalne zmiany średnich wartości drugiej cechy (wzrost lub spadek).




**Kowariancja** (*cov*; średnia arytmetyczna iloczynu odchyleń wartości zmiennych X i Y od ich średnich arytmetycznych) **jest miarą związku liniowego pomiędzy dwoma zmiennymi**.

$$cov(x, y) = \frac{\sum(x_i - x_{sr})(y_i - y_{sr})}{n}$$

Z uwagi na to, że wartość kowariancji zależy od rzędu wielkości w jakich wyrażone są obydwie zmienne, nie można jej wykorzystać w sposób bezpośredni do porównań (można ją określić jako miarę pośrednią).

**Współczynnik korelacji Pearsona *r*** (wyznaczany przez standaryzację kowariancji) **jest już miarą unormowaną** (przyjmuje wartości od -1 do 1).

$$r_{xy} = \frac{cov(x, y)}{Sd_x \times Sd_y}$$

# 3.1. Współczynnik korelacji Pearsona.

**Współczynnik korelacji Pearsona** -> unormowany miernik natężenia i kierunku współzależności liniowej dwóch zmiennych mierzalnych *X* i *Y*.

Założenia: obie zmienne powinny być zmiennymi ciągłymi z normalnym rozkładem.

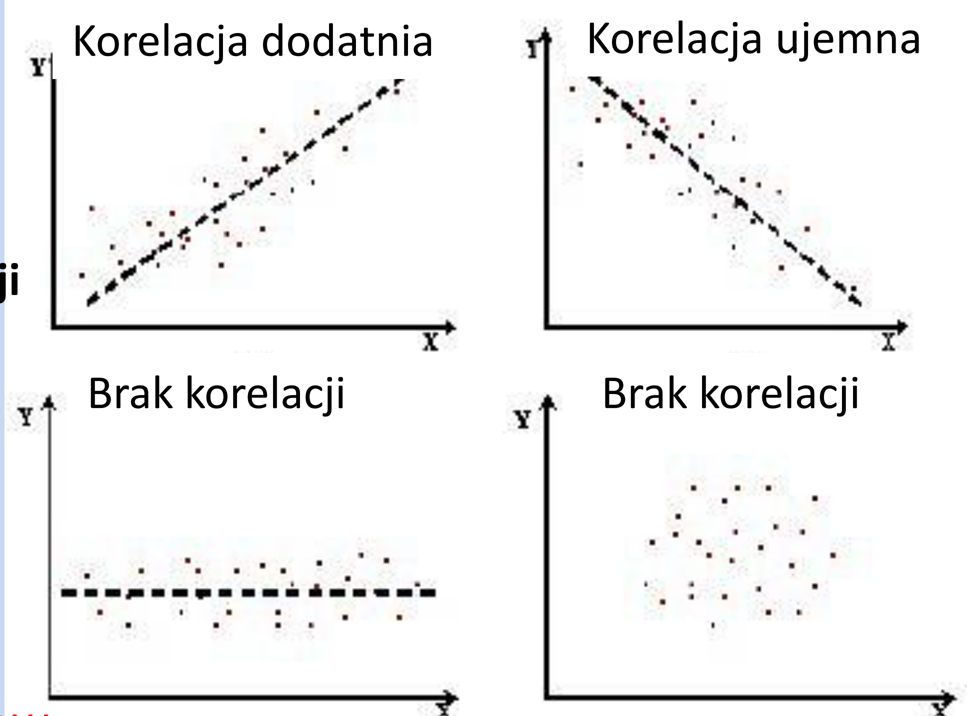
Charakterystyka: może przyjmować wartości od -1 do 1.

**Znak** wskazuje na **kierunek zależności** (wraz ze wzrostem wartości jednej zmiennej odpowiednio wzrastają bądź maleją wartości drugiej).

- wartość współczynnika korelacji od 0 do 1 informuje, że wzrostowi wartości jednej cechy towarzyszy wzrost średnich wartości drugiej cechy (korelacja dodatnia),
- wartość współczynnika korelacji od -1 do 0 informuje, że wzrostowi wartości jednej cechy towarzyszy spadek średnich wartości drugiej cechy (korelacja ujemna).

**Wartość bezwzględna** współczynnika korelacji wskazuje na **siłę istniejącej korelacji** -> im większa tym korelacja jest silniejsza.

- wartość -1 oznacza występowanie doskonałej korelacji ujemnej (punkty leżą dokładnie na prostej, skierowanej w dół),
- wartość 1 oznacza doskonałą korelację dodatnią (punkty leżą dokładnie na prostej, skierowanej w górę).
- wartość 0 oznacza brak korelacji liniowej.




Khelil i in. 2007

Uwaga: najważniejsza jest istotność korelacji!!!

# 4. Funkcje do obliczania korelacji w R.

- ogólna postać miary korelacji:

```
> cor( var1, var2, method = "method")
```

- opcja domyślna to miara korelacji Pearsona:

```
> cor(var1, var2)
```

- gdy chcemy użyć zbioru danych zamiast osobnych zmiennych:

```
> cor(dataset, method = "pearson")
```

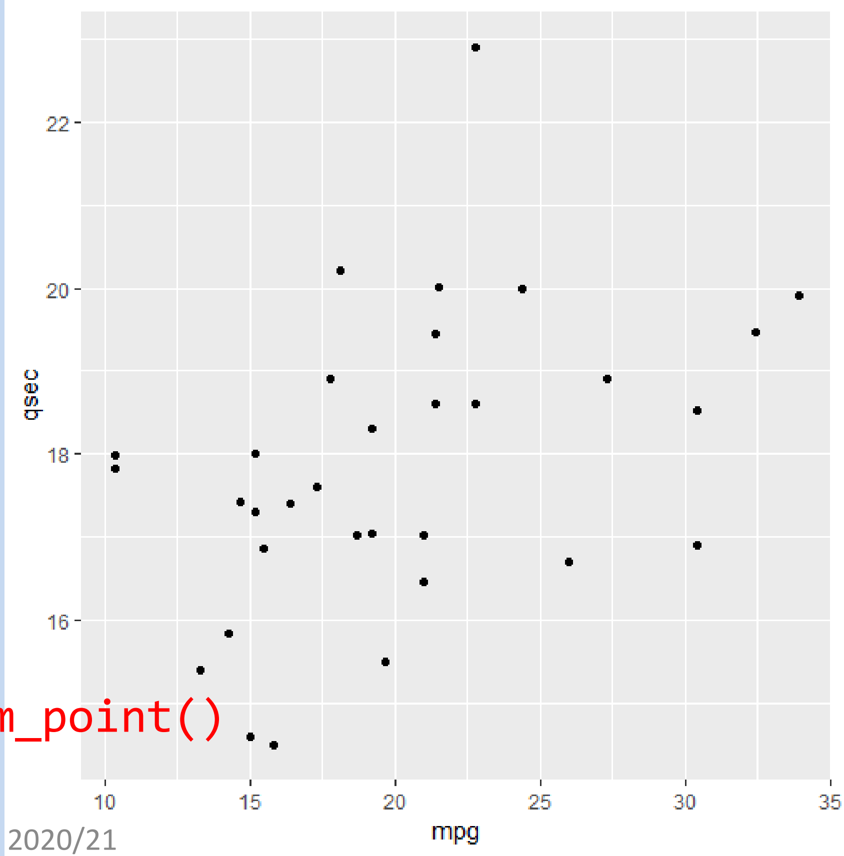
Zadanie: Oblicz współczynnik korelacji dla zmiennych mpg i wt z bazy mtcarts.

```
> cor( mtcars$mpg, mtcars$wt)
[1] -0.8676594
```

Zadanie: Zbadaj związek pomiędzy zmiennymi mpg i qsec z bazy mtcars

```
> cor( mtcars$mpg, mtcars$qsec)
[1] 0.418684
```

```
> ggplot(mtcars, aes(mpg, qsec))+geom_point()
```



# 4. Funkcje do obliczania korelacji w R.

Wykorzystamy dane gridowe z modelu CRU TS v. 1.2 (średnie miesięczne temperatury i miesięczne sumy opadów). Więcej informacji na temat danych tutaj: https://crudata.uea.ac.uk/cru/data/hrg/).

Skorzystamy z danych dla dwóch „oczek” siatki (dane w siatce 10’x10’) dla dwóch lokalizacji w Karpatach: dla okolic Cisnej (ci2_CRU2_1) i okolic Pcimia (my1_CRU2_1). Dane dla temperatury na końcu nazwy mają „t”, dla opadów – „p”.

Ściągnij dane z MS Teams z katalogu cwiczenia 10/dane. Zaimportuj dane do Rstudio.

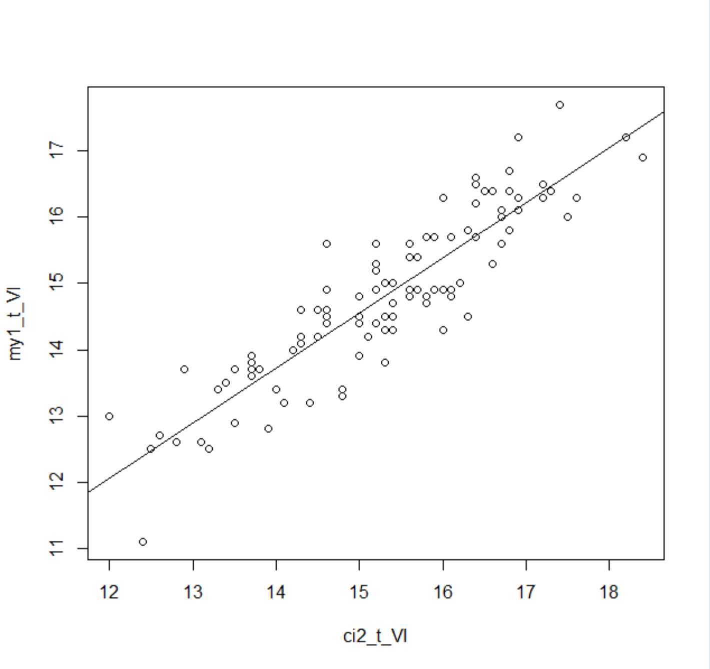
# 4. Funkcje do obliczania korelacji w R.

Zadanie 1: Zbadaj związek (wykorzystując współczynnik korelacji Pearsona) pomiędzy średnią temperaturą w czerwcu dla wcześniej wspomnianych lokalizacji:

```
> cor(ci2_CRU2_1_t$VI, my1_CRU2_1_t$VI)
[1] 0.8983227
```

Wartość współczynnika jest wysoka, ale pamiętajmy  jeszcze o narysowaniu wykresu zależności:

```
> plot(ci2_CRU2_1_t$VI, my1_CRU2_1_t$VI, xlab="ci2_t_VI", ylab="my1_t_VI")
```




Wyraźnie widać liniowy charakter związku.

Na wykresie dodajmy jeszcze (funkcją abline) linię trendu (czyli prostą regresji – funkcja `lm` służy do budowy modelu regresji liniowej – więcej na ten temat na jednych z kolejnych zajęć):

```
> abline(lm(my1_CRU2_1_t$VI~ci2_CRU2_1_t$VI))
```

Uwaga: Zauważ, że w powyższej formule kolejność podanych argumentów dla funkcji `lm` jest odwrotna niż w poleceniu `plot` -> prosta regresji wyznaczana jest na podstawie modelu, w którym „y” chcemy szacować na podstawie „x”.

# 4. Funkcje do obliczania korelacji w R.

Zadanie 2. Narysuj wykres rozrzutu z zadania 2, ale w `ggplot2`.

Podpowiedź: Najpierw musimy utworzyć nową ramkę danych z wykorzystaniem funkcji `select` i `left_join` (lub `select` i `data.frame`)

# 4. Funkcje do obliczania korelacji w R.

Zadanie 2 – odpowiedź:

Wybieramy co trzeba z istniejących ramek (`select`):

```
> ci2_t_VI<-select(ci2_CRU2_1_t, rok, ci2_t_VI=VI)
> my1_t_VI<-select(my1_CRU2_1_t, rok, my1_t_VI=VI)
```

I tworzymy nową ramkę danych (`left_join`):

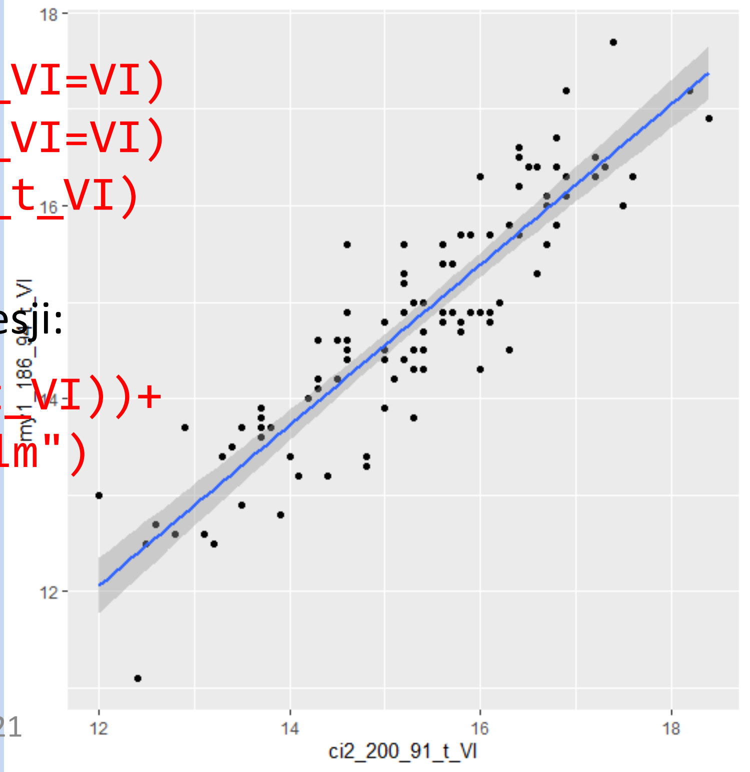
```
> t_VI_my1_ci2<-left_join(ci2_t_VI, my1_t_VI, by="rok")
```

Innym sposobem, przy użyciu `select` i `data.frame` nie musimy dodawać kolumny łączącej):

```
> my1_t_VI <- select(my1_CRU2_1_t, my1_t_VI=VI)
> ci2_t_VI <- select(ci2_CRU2_1_t, ci2_t_VI=VI)
> VI_t_my1_ci2<-data.frame(my1_t_VI, ci2_t_VI)
```

Wreszcie rysujemy wykres rozrzutu dodając prostą regresji:

```
> ggplot(ci2_my1_t_V, aes(ci2_t_VI, my1_t_VI))+
        geom_point()+geom_smooth(method="lm")
```



# 4. Funkcje do obliczania korelacji w R.

Zadanie 3: Wykonaj analizę jak w zadaniu 1, ale dla opadów.

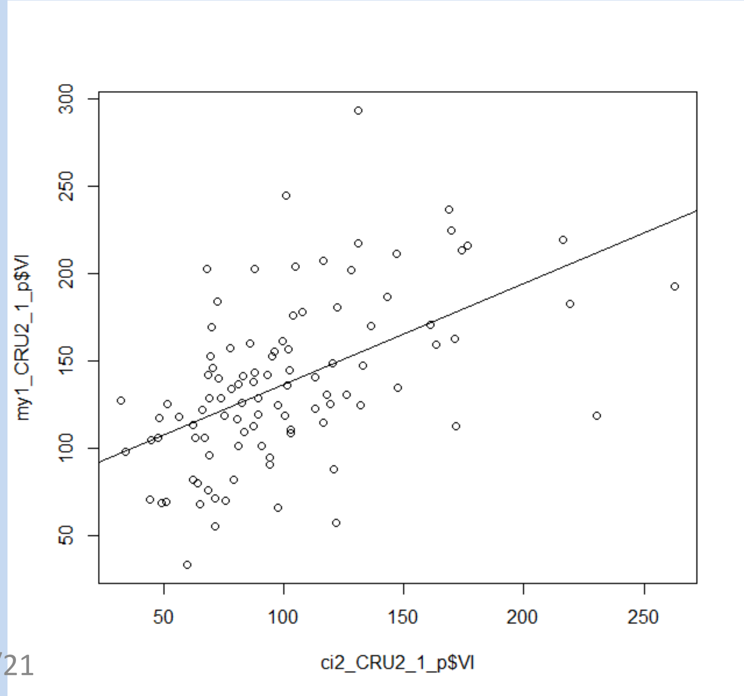
```
> cor(ci2_CRU2_1_p$VI, my1_CRU2_1_p$VI)
[1] 0.5245463
```

Wartość współczynnika jest znacznie niższa (nadal jednak wysoka), sprawdźmy wykres:

```
> plot(ci2_CRU2_1_p$VI, my1_CRU2_1_p$VI)
```

Związek liniowy jest słabiej widoczny, jednak nadal możemy się go dopatrzeć.

```
> abline(lm(my1_CRU2_1_p$VI~ci2_CRU2_1_p$VI))
```




Wyniki analizy zwracają uwagę na wyraźnie większą lokalność opadów w porównaniu z temperaturą.

Uwaga: Koniecznie należy także sprawdzić istotność korelacji!!!

patrz 2 slajdy dalej

# 4. Funkcje do obliczania korelacji w R.

Zadanie 3: Oblicz współczynnik korelacji Pearsona dla temperatury i opadów w maju, dla stacji  Karpacz (dane są do pobrania w katalogu cwiczenia 10/dane na MS Teams).

```
> cor(karpacz_t$V, karpacz_p$V)
[1] NA
```

Coś jest nie tak. Spójrzmy na dane:

```
> View(karpacz_t)
> View(karpacz_p)
```

Funkcja `cor` działa domyślnie tylko dla pełnych danych. Rozwiązaniem jest albo uzupełnienie danych lub też ich ominięcie w procesie jej obliczania.
Pomińmy luki w danych:

```
> cor(karpacz_t$V, karpacz_p$V, use="complete.obs")
[1] -0.3664717
```

# 3.1. Istotność korelacji

Uwaga: Związek może być silny (wysoka wartość współczynnika korelacji), a mimo to nieistotny i odwrotnie – słaby ale istotny.
Kluczowa jest tutaj wielkość próby.

- Dla małych zbiorów danych jest stosunkowo łatwo uzyskać silną korelację przez przypadek **i trzeba zwrócić uwagę na poziom istotności** zanim wyciągnie się ostateczne wnioski (aby nie odrzucić prawdziwej hipotezy zerowej, czyli nie popełnić błędu I rodzaju).

- Dla większych zbiorów jest bardzo łatwo osiągnąć istotność, ale trzeba zwrócić uwagę na **siłę korelacji** (wartość bezwzględną współczynnika korelacji), aby mieć pewność, że mamy do czynienia z rzeczywistym związkiem.

# 4. Funkcja do obliczania istotność korelacji w R

Jeśli chcemy poznać stopień istotności korelacji między badanymi zmiennymi musimy użyć dodatkowo funkcji do testowania korelacji `cor.test()`, domyślnie stosowana jest tutaj korelacja Pearsona.

Obliczmy istotność statystyczną korelacji dla zmiennych `mpg` i `wt` z bazy `mtcarts`:

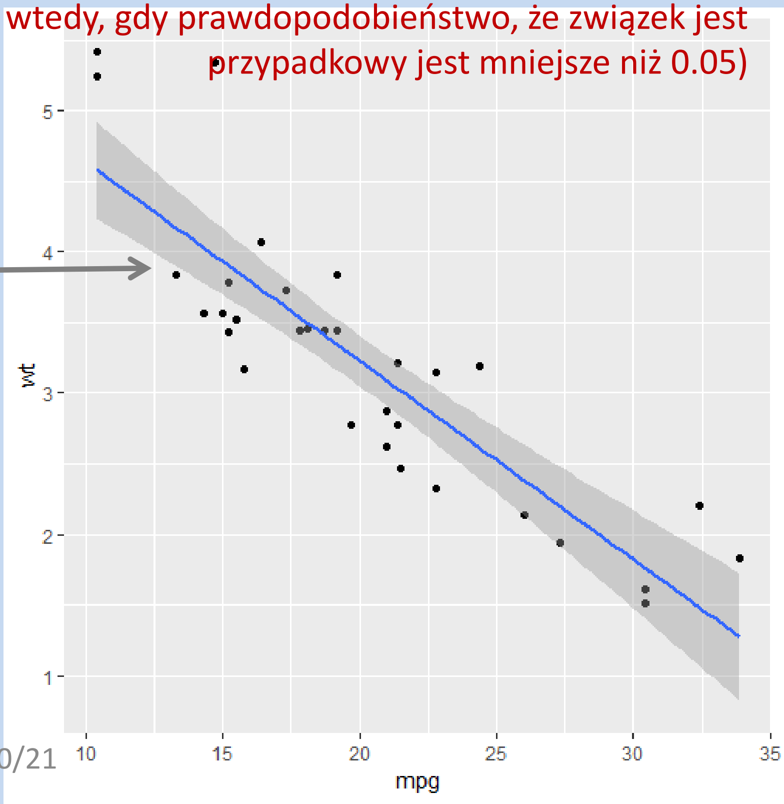
```
> cor.test( mtcars$mpg, mtcars$wt)

Pearson's product-moment correlation
data:  mtcars$mpg and mtcars$wt
t = -9.559, df = 30, p-value = 1.294e-10
alternative hypothesis: true correlation is
not equal to 0
95 percent confidence interval:
 -0.9338264 -0.7440872
sample estimates:
       cor
-0.8676594
```

im mniejsza wartość *p* tym większe prawdopodobieństwo, że korelacja jest prawdziwa; przyjmijmy, że korelacja jest istotna, gdy $p<0.05$ (czyli wtedy, gdy prawdopodobieństwo, że związek jest przypadkowy jest mniejsze niż 0.05)

na 95% wartość korelacji mieści się w tym przedziale

Narysujmy jeszcze raz wykres wraz z prostą regresji:

```
> ggplot(mtcars, aes(mpg, wt))+
        geom_point()+
        geom_smooth(method="lm")
```



# 4. Funkcja do obliczania istotność korelacji w R

Zadanie: Zbadaj związek pomiędzy temperaturą w lipcu dla stacji w Karpaczu i Jeleniej Górze. Czy korelacja ta jest istotna statystycznie? Jak zinterpretujesz wyniki?
Uwaga: Pamiętaj o lukach w danych ze stacji Karpacz.

```
> cor.test (karpacz_t$VII, jelenia_gora_t$VII, method="pearson",
use="complete.obs")

Błąd w poleceniu 'cor.test.default(karpacz_t$VII, jelenia_gora_t$VII,
method = "pearson", ':
  argumenty 'x' oraz 'y' muszą mieć tę samą długość
```

# 4. Funkcja do obliczania istotność korelacji w R

Zobaczmy jak wyglądają dane (`View`). Dane są różnej długości, aby wykonać analizę korelacji dane muszą być tej samej długości.
Musimy więc „przyciąć” o jeden rok (ostatni) dane ze stacji w Jeleniej Górze.

```
> jelenia_gora_t_1<-jelenia_gora_t[1:64,]
> cor.test (karpacz_t$VI, jelenia_gora_t_1$VI, method="pearson",
use="complete.obs")
```

```
Pearson's product-moment correlation
data:  karpacz_t$VI and jelenia_gora_t_1$VI
t = 16.131, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8395000 0.9385264
sample estimates:
      cor
0.9000548
```

Czy pamiętaliście o narysowaniu wykresu rozrzutu (korelogramu)?
Jakie są wnioski?

# 3.3. Współczynnik korelacji rang Spearmana

**Współczynnik korelacji rang Spearmana** wykorzystywany jest do opisu siły korelacji dwóch cech, w przypadku gdy:

- zmienne mają charakter jakościowy, pozwalający na uporządkowanie ze względu na siłę tej cechy,
- zmienne mają charakter ilościowy, ale ich liczebność jest niewielka,
- rozkład zmiennych nie jest normalny (np. w wyniku występowania wartości odstających),
- zależność zmiennych nie jest liniowa(ale jednak jest monotoniczna; na wykresie rozproszenia punkty rozproszone są np. wokół krzywej logarytmicznej, wykładniczej, itp.)

$$r_s = 1 - \frac{6\sum_{i=0}^{n} d_i^2}{n(n^2-1)}$$

gdzie:
$d_i^2$ – różnica pomiędzy rangami odpowiadających sobie wartości zmiennych $x_1$ i $y_i$.

# 3.3. Współczynnik korelacji rang Spearmana

**Ranga** jest to liczba odpowiadająca miejscu w uporządkowaniu każdej ze zmiennych.

| osobnik | cecha x | cecha y | Ranga x | Ranga y | $d_i$ (różnica między rangami) | $d_i^2$ |
|---|---|---|---|---|---|---|
| A | 42 | 39 | 7 | 6.5 | 0.5 | 0.25 |
| B | 27 | 24 | 2 | 1 | 1 | 1 |
| C | 36 | 35 | 4 | 5 | -1 | 1 |
| D | 33 | 29 | 3 | 3 | 0 | 0 |
| E | 24 | 26 | 1 | 2 | -1 | 1 |
| F | 47 | 47 | 9 | 9 | 0 | 0 |
| G | 39 | 44 | 6 | 8 | -2 | 4 |
| H | 52 | 51 | 10 | 10 | 0 | 0 |
| I | 43 | 39 | 8 | 6.5 | 1.5 | 2.25 |
| J | 37 | 32 | 5 | 4 | 1 | 1 |
| suma: | | | | | | 10.5 |

najniższa wartość cechy – ranga 1

najwyższa wartość cechy – ranga: 10

jeśli w badanej zbiorowości jest więcej jednostek z identycznym natężeniem badanej zmiennej (tu: dla dwóch osobników wartość *cechy y* to 39), to jednostkom tym przypisuje się identyczne rangi, licząc średnią arytmetyczną z rang przynależnych tym samym jednostkom (czyli nie mamy rangi 6 ani 7 ale 6.5 dla obydwu wartości 39).

Współczynnik korelacji Spearmana dla cech *x* i *y*:

$$r_s = 1 - \frac{6\sum_{i=0}^{n} d_i^2}{n(n^2-1)} \quad \Rightarrow \quad r_s = 1 - \frac{6 \times 10.5}{10(100-1)} = \frac{63}{990} = 0.936$$

 Przykład zaczerpnięty z: http://zsi.tech.us.edu.pl/~nowak/odzw/korelacje.pdf

# 3.3. Współczynnik korelacji rang Spearmana

- współczynnik korelacji rang przyjmuje wartości z przedziału [-1; 1]
- interpretacja jest podobna do współczynnika korelacji liniowej Pearsona…
- … jednak korelacja rangowa

  **- pokazuje dowolną monotoniczną zależność** (także nieliniową)

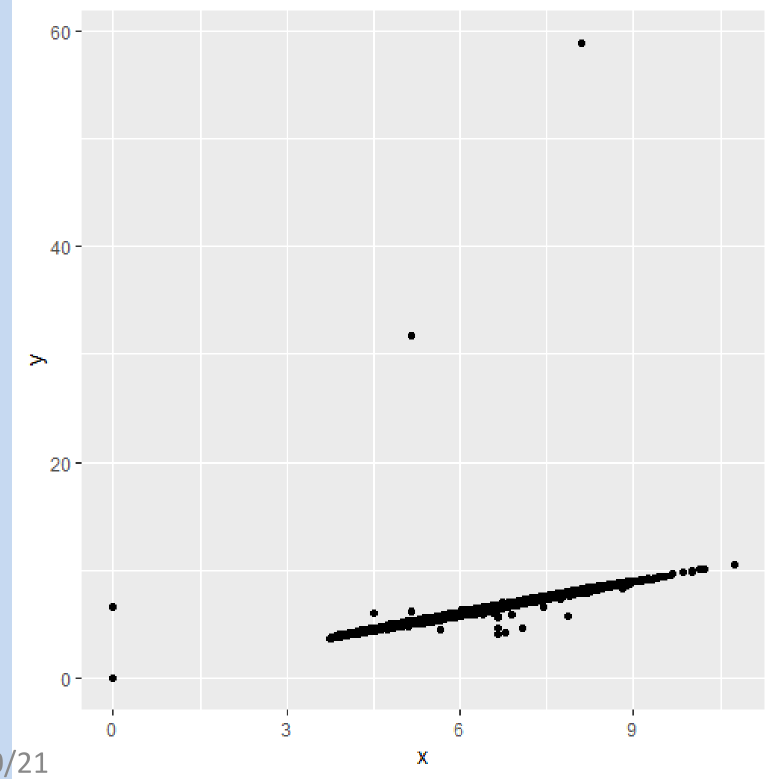
  - **jest w niewielkim tylko stopniu wrażliwa na obserwacje odstające** (ang. *outliers*), dzięki czemu szczególną użyteczność znajduje w analizie danych niskiej jakości.

# 3.3. Współczynnik korelacji rang Spearmana

Zadanie: zbadaj związek pomiędzy zmiennymi *x* i *y* diamentów z bazy `diamonds` z wykorzystaniem współczynników korelacji Pearsona i rang Spearmana.

```
> ggplot (diamonds, aes(x, y)) + geom_point()
> cor(diamonds$x, diamonds$y)
[1] 0.9747015
> cor(diamonds$x, diamonds$y, method="spearman")
[1] 0.9978949
```

Co mogło być powodem wyższej wartości otrzymanej dla współczynnika rang Spearmana?*




* Przykład ten znowu pokazuje, że przy badaniu związków między zmiennymi warto popatrzeć na wykres rozrzutu. Występuje tutaj kilka wartości wyraźnie odstających od reszty, które prawdopodobnie wpłynęły na niższą wartość współczynnika korelacji Pearsona w porównaniu z wartością współczynnika korelacji rang Spearmana (który jest bardziej „odporny” w tego typu sytuacjach).

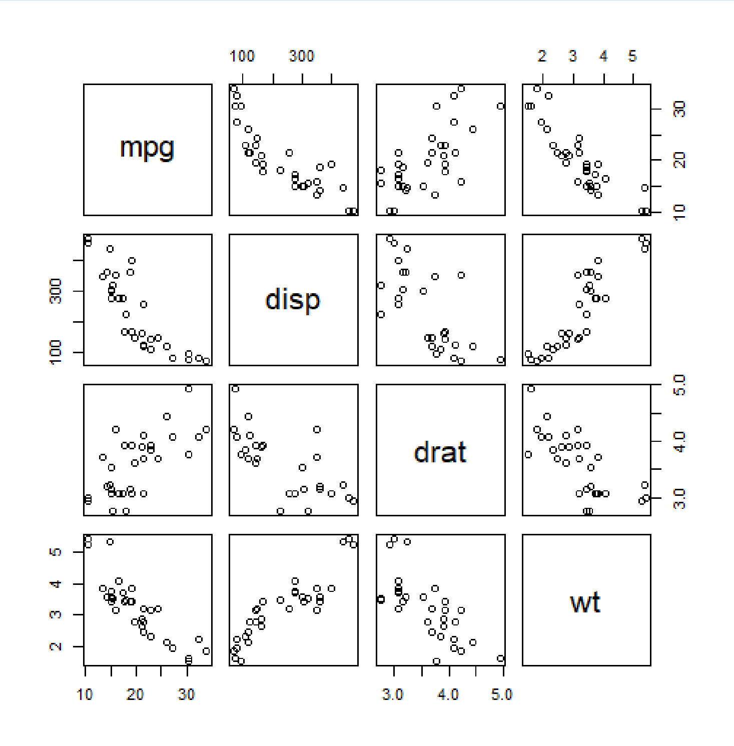
# 4. Reprezentacja graficzna korelacji w R

- wykresy rozrzutu (inaczej korelaogramy),
- wykresy rozrzutu (korelogramy) dla wielu zmiennych.

# 4.1. Macierze wykresów rozproszenia.

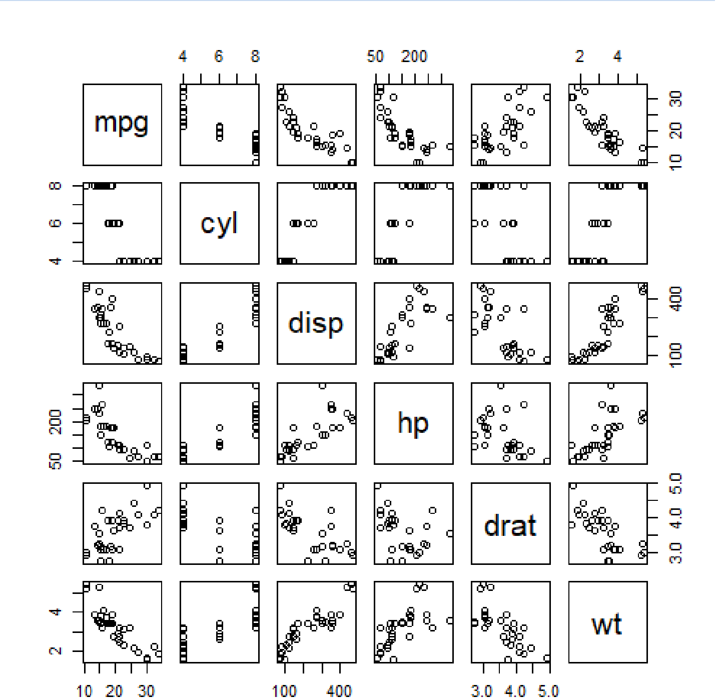
Skorzystamy z funkcji `pairs` z pakietu `base`.
Narysujmy wykresy rozproszenia (w parach)
dla zmiennych `mpg`, `disp`, `drat` oraz `wt`
(baza danych `mtcars`):

```
> pairs(~mpg+disp+drat+wt, data=mtcars)
```




Sporządźmy podobny wykres dla pierwszych 6 zmiennych:

```
> pairs(mtcars [1:6])
```



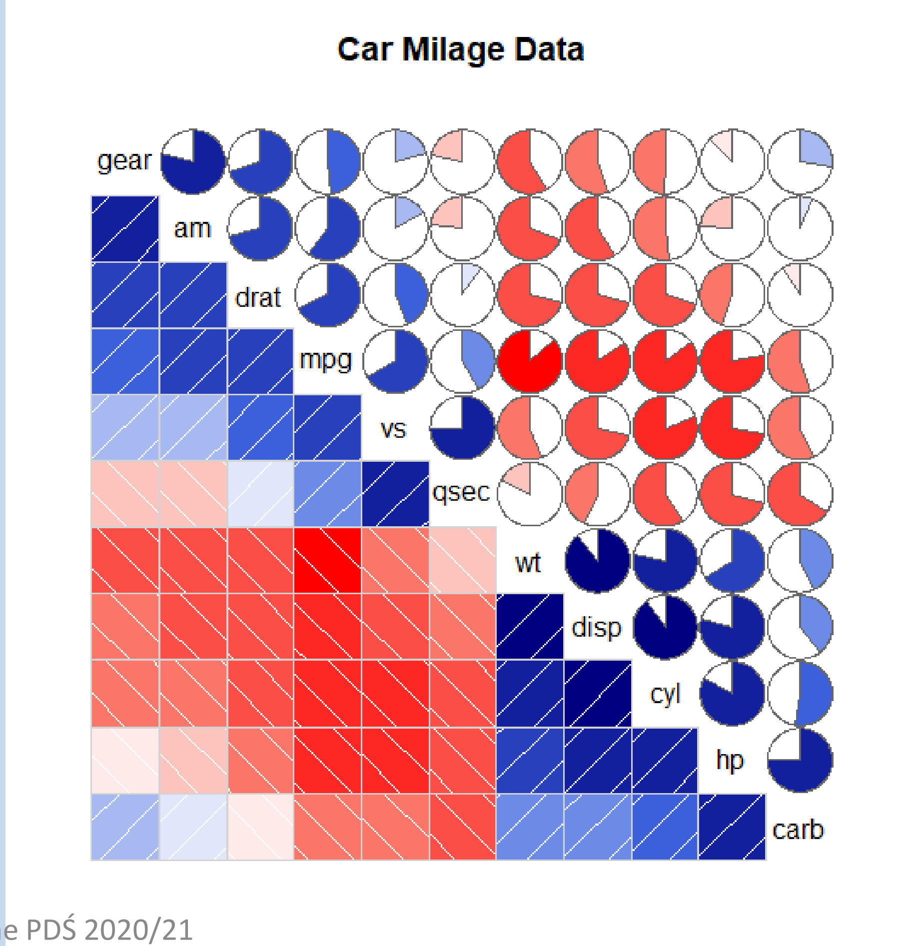
# 4.2. Korelogramy dla macierzy korelacyjnych.

Pakiet i funkcja `corrgram` pozwala na wizualizację macierzy korelacyjnych, czyli tabel przedstawiających współczynniki korelacji między poszczególnymi zmiennymi (w parach).

```
> install.packages("corrgram")
> library(corrgram)
```

Zobaczmy jak wyglądają korelacje pomiędzy zmiennymi bazy `mtcars`:

```
> corrgram(mtcars,
order=TRUE,
lower.panel=panel.shade,
upper.panel=panel.pie,
text.panel=panel.txt,
main= "Car Milage Data")
```




*kolorem czerwonym są zaznaczone korelacje dodatnie, niebieskim – ujemne; odcienie czerwonego i niebieskiego pokazują „siłę" korelacji (czyli wartość)

# Komentarz

To, że dwie (lub więcej) zmienne są silnie ze sobą skorelowane nie musi oznaczać, że jedna zmienna jest zależna od drugiej.

Więcej na ten temat np. tutaj:

- https://pl.wikipedia.org/wiki/Zale%C5%BCno%C5%9B%C4%87_zmiennych_losowych#Zale%C5%BCno%C5%9B%C4%87_statystyczna
  (w sekcji: zależność a wynikanie)

„Korelacja nie dowodzi (…) żadnego związku przyczynowo-skutkowego”

- https://algolytics.pl/korelacja-to-nie-to-samo-co-zaleznosc/

„Korelacja oznacza, że znając wartość jednej zmiennej możemy wnioskować na temat wartości drugiej”

- http://www.ekonomiawprzykladach.pl/2014/05/korelacja-przyczynowosc.html

- http://www.tylervigen.com/spurious-correlations
  (przykłady korelacji pozornych)

# Wykorzystane materiały/tu możesz znaleźć więcej:

- https://www.statsoft.pl/textbook/stathome.html
- geo.uni.lodz.pl/uploads/jandegir/STAT_wyklad_6.pdf
- zsi.tech.us.edu.pl/~nowak/odzw/SMAD_korelacje.pdf
- zsi.tech.us.edu.pl/~nowak/odzw/korelacje.pdf