



BSS
BIAŁOWIESKA SZKOŁA STATYSTYKI

Tworzenie wykresów przy użyciu pakietu ggplot2



mgr Patryk Czortek, mgr inż. Marcin K. Dyderski

Pakiet ggplot2

- jedno z najbardziej zaawansowanych narzędzi do tworzenia wykresów statystycznych
- Zaawansowanie nie oznacza, że można szybko zrobić w nim wykres, ani też, że dostępnych jest wiele szablonów wykresów
- Oznacza, że konstrukcja pakietu jest na tyle elastyczna, że można z nim wykonać praktycznie każdą grafikę statystyczną

Wykres punktowy

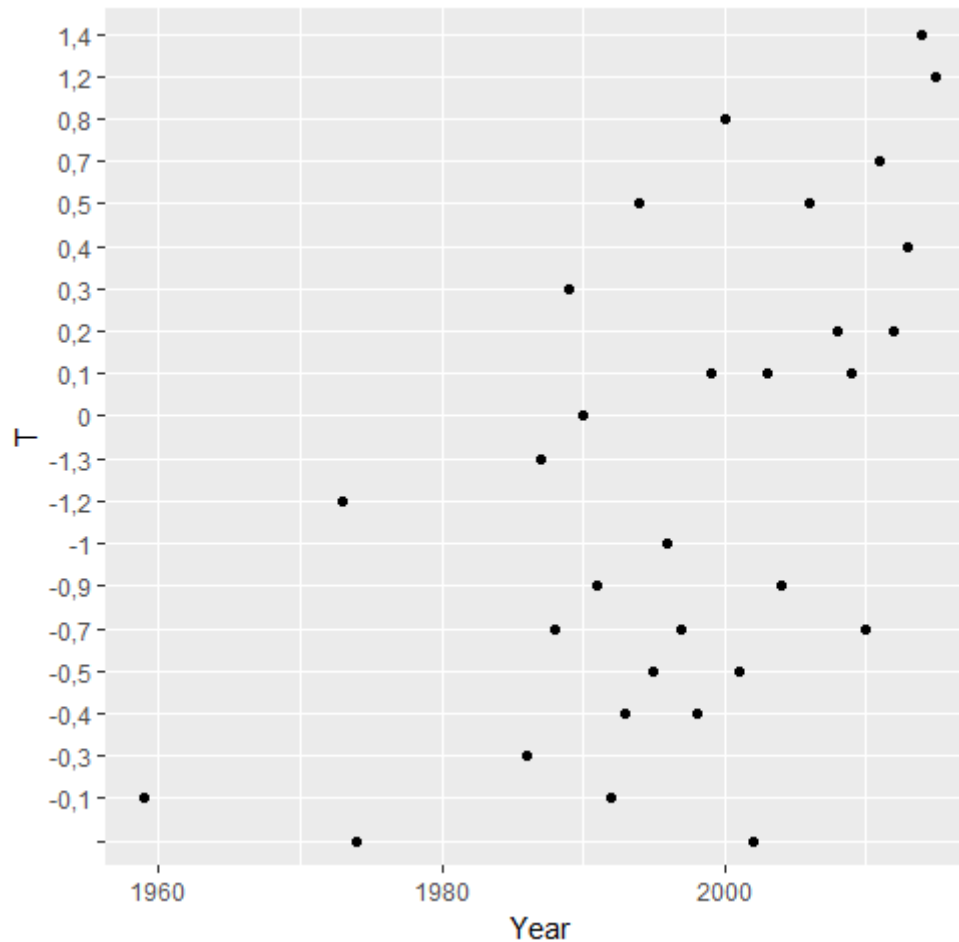
#załadowanie danych do R

```
climate.kasprowy<-read.csv('climate.kasprowy.csv', sep=";")
```

	Year	T	TM	Tm	PP	V	RA	SN	TS	FG	TN	GR
1	1959	-0,1	2,5	-2,7		18,2	97	139	35	211	0	2
2	1973	-1,2	1,4			27	80	138	16	220	0	0
3	1974						87	150	19	249	1	5
4	1986	-0,3	1,9	-2,7		24,8	68	111	21	224	0	1
5	1987	-1,3	1	-3,8	1335,27	26,7	85	138	16	237	0	1
6	1988	-0,7	1,9	-3,3	1551,91	25,7	72	141	22	242	0	2
7	1989	0,3	3,1	-2,4	1588,86	25,4	95	122	29	245	0	5
8	1990	0	2,7	-2,5	1394,98	26,1	88	143	26	254	0	2
9	1991	-0,9	1,8	-3,2	2004,62	22,7	94	141	16	266	0	3
10	1992	-0,1	3	-2,8	1681,51	23,1	73	134	24	237	0	0
11	1993	-0,4	2,6	-3,1	1262,68	23	89	121	24	231	0	0
12	1994	0,5	3,3	-2,2	1647,52	23	77	139	22	244	0	1
13	1995	-0,5	2,2	-3	1683,58	22,4	72	155	29	249	0	1
14	1996	-1	1,8	-3,5	1859,58	20,8	100	142	31	237	0	1
15	1997	-0,7	2,1	-3,3	1987,1	22,1	83	141	28	261	1	3
16	1998	-0,4	2,5	-3	1883,93	24	103	137	20	237	0	0
17	1999	0,1	3,1	-2,8	1605,72	22,6	120	146	28	267	0	7
18	2000	0,8	4	-2,4	1656,08	23,7	125	141	35	257	0	6
19	2001	-0,5	2,4	-3,5		23,3	125	161	26	278	0	7
20	2002						120	128	44	248	0	3

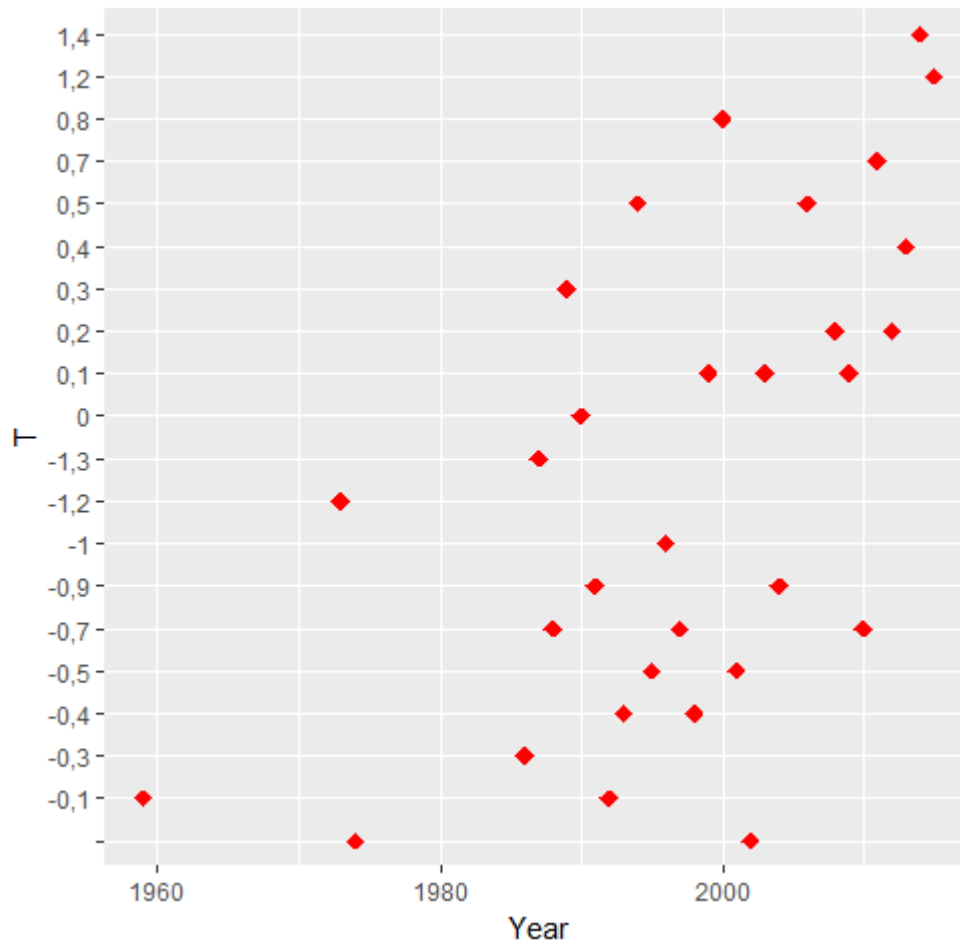
#tworzenie wykresu

```
scatter.plot<-ggplot(climate.kasprowy, aes(x=Year, y=T)) +  
  geom_point()
```



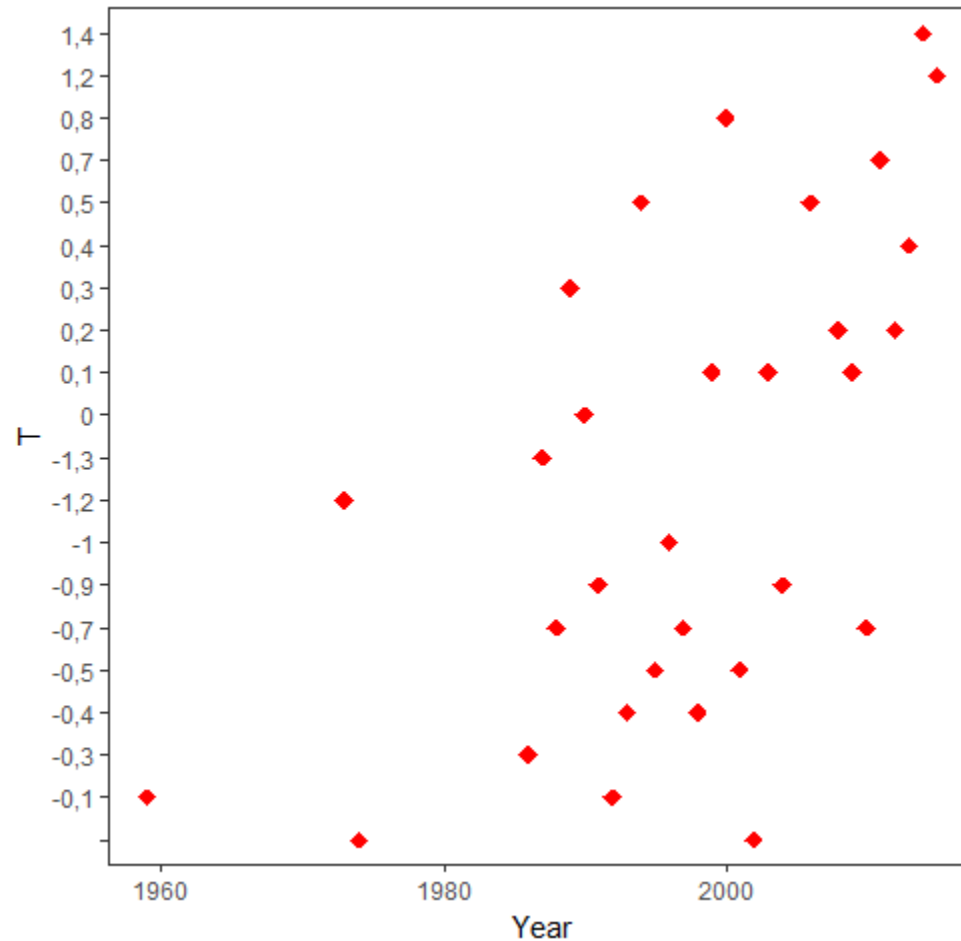
#zmiana kształtu, rozmiaru i koloru punktów

```
scatter.plot+geom_point(shape=18, size=3, color="red")
```



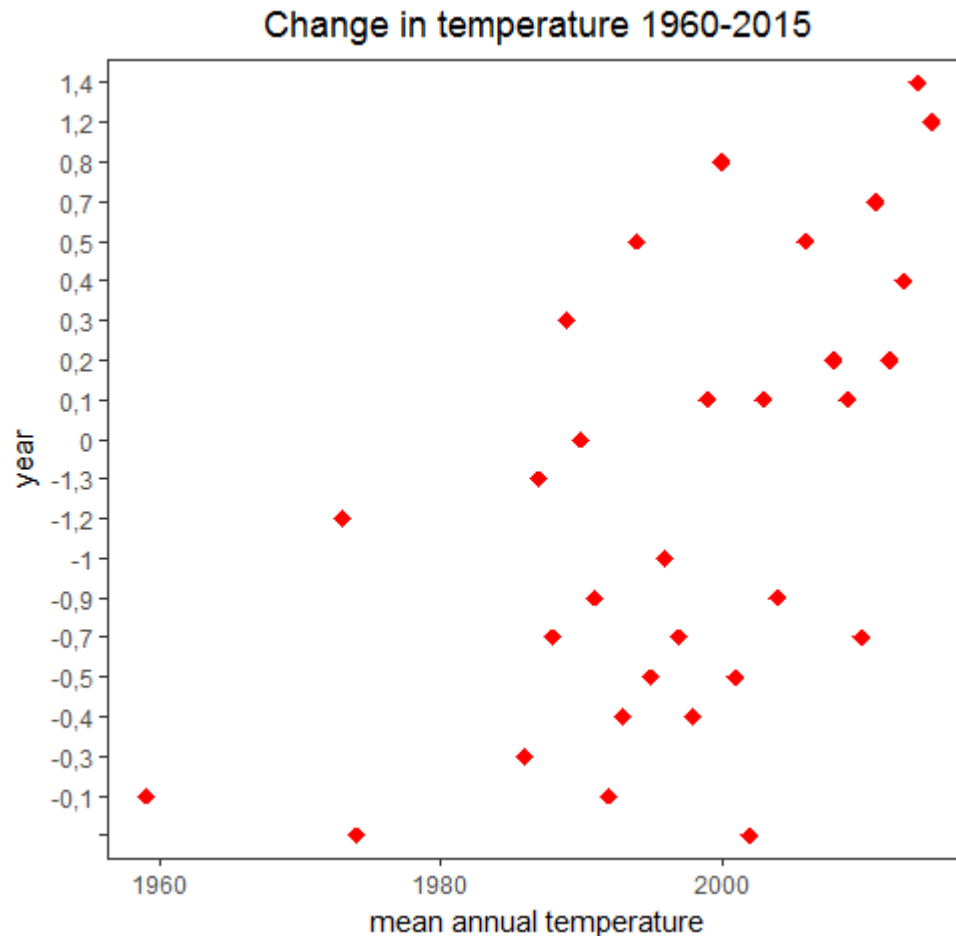
#usuwanie tła i podziałki

```
scatter.plot+theme_bw()+theme(panel.grid =  
element_blank())
```



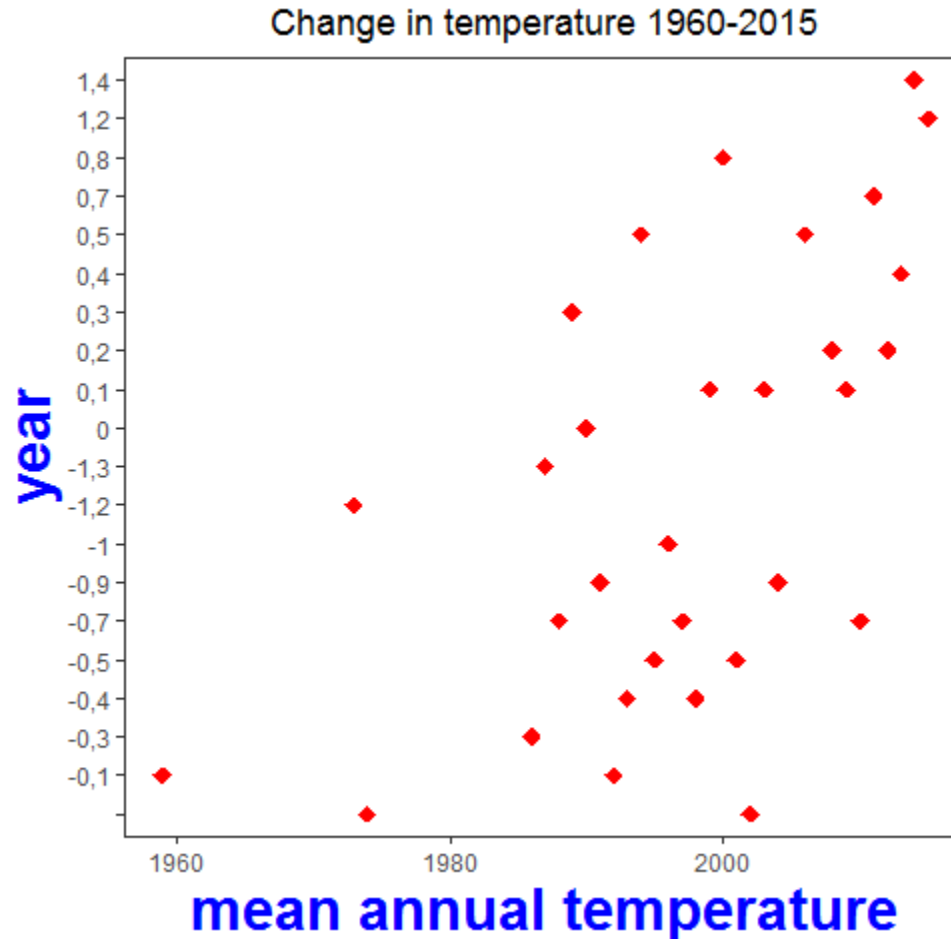
#zmiana nazw osi i dodanie tytułu wykresu

```
scatter.plot+labs(x = "mean annual temperature", y="year",  
size=12)+ggtitle("Change in temperature 1960-  
2015")+theme(plot.title = element_text(hjust = 0.5))
```



#zmiana koloru i rozmiaru oraz pogrubienie nazw osi

```
scatter.plot+theme(axis.title = element_text(color="blue",  
face="bold", size=22))
```



Wykres pudełkowy

#załadowanie danych do R

```
traity.hab<-read.csv('traity.hab.csv', sep=";")
```

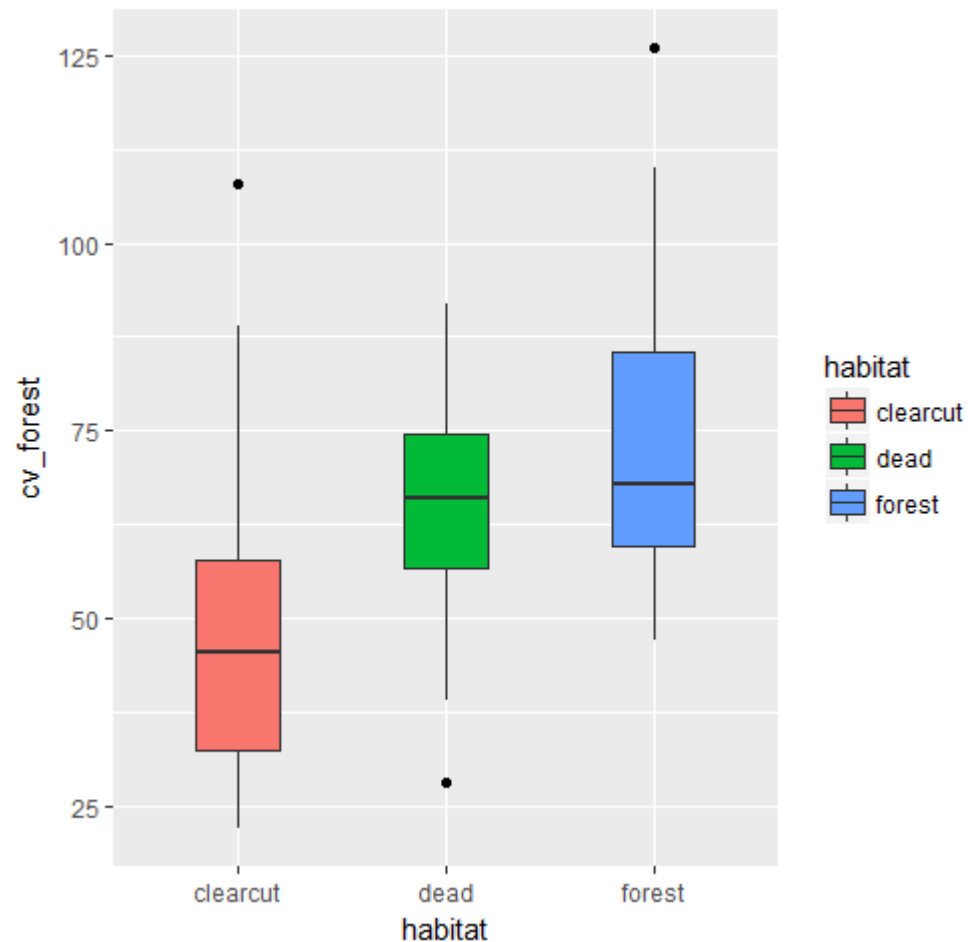
	EIV_L	EIV_T	EIV_F	EIV_N	cv_open	cv_forest	shan	even	rich
c_1	5.263158	5.000000	5.714286	6.490196	32	25	3.376736	0.9217084	39
c_10	6.000000	5.010638	4.761905	4.939394	58	55	2.715321	0.7464622	38
c_11	4.922222	4.980392	5.160494	5.674419	23	43	2.944761	0.8095370	38
c_12	4.917526	5.050000	5.096154	3.978495	35	56	2.699769	0.7861908	31
c_13	4.833333	5.648649	5.039062	5.000000	9	108	2.206828	0.6311520	33
c_14	6.312500	5.023256	6.394958	5.148760	76	25	2.994279	0.7820736	46
c_15	5.703704	4.854545	5.344262	6.031746	48	50	3.168567	0.8373172	44
c_16	4.615385	5.000000	6.064516	4.305556	16	22	2.712203	0.8650001	23
c_17	4.605263	4.982143	5.112676	5.400000	10	57	2.547015	0.7349131	32
c_18	3.949153	4.800000	5.169811	6.142857	15	36	2.846534	0.8369211	30
c_19	4.819444	4.925000	5.819672	5.238095	15	42	2.920524	0.8764541	28
c_2	5.260000	5.266667	5.764706	5.903846	23	30	3.073795	0.8869098	32
c_20	4.421875	5.780488	5.166667	5.055556	5	55	2.233195	0.7026927	24
c_21	3.968750	5.200000	5.360000	5.961538	3	26	2.740951	0.9002895	21
c_22	4.378378	5.233333	5.588235	5.969697	6	32	2.779671	0.8992665	22

#tworzenie wykresu

```
dodge <- position_dodge(width = 0.4)
```

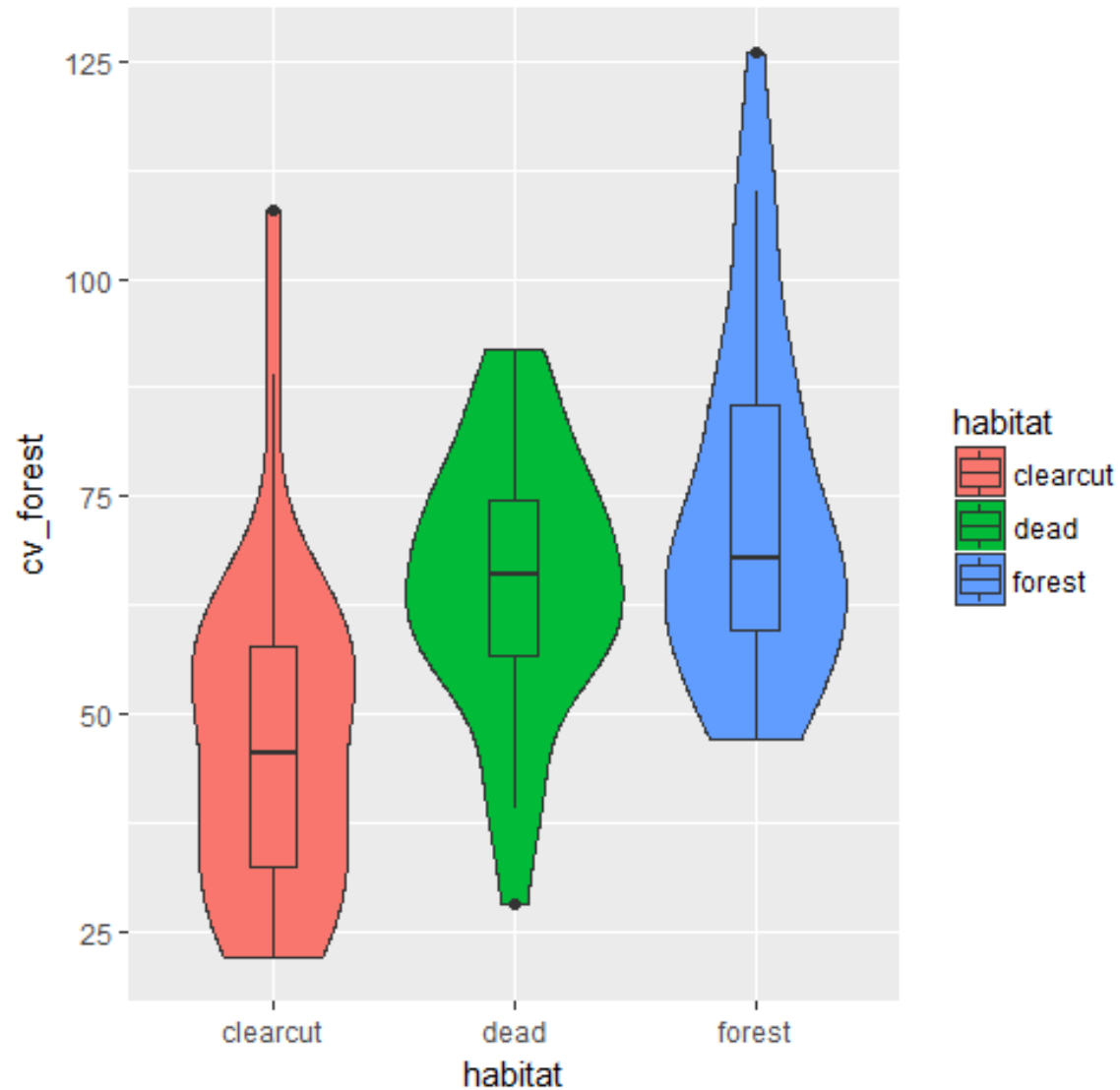
```
box.plot<-ggplot(data = traity.hab, aes(x = habitat, y =  
cv_forest, fill = habitat)) +
```

```
  geom_boxplot(data=traity.hab, mapping=aes(x=habitat, y=  
cv_forest), width=0.4, outlier.color="black")
```



#wykres skrzypcowy

```
box.plot+ geom_violin()+ geom_boxplot(width=0.2)
```



Kilka wykresów na jednej stronie

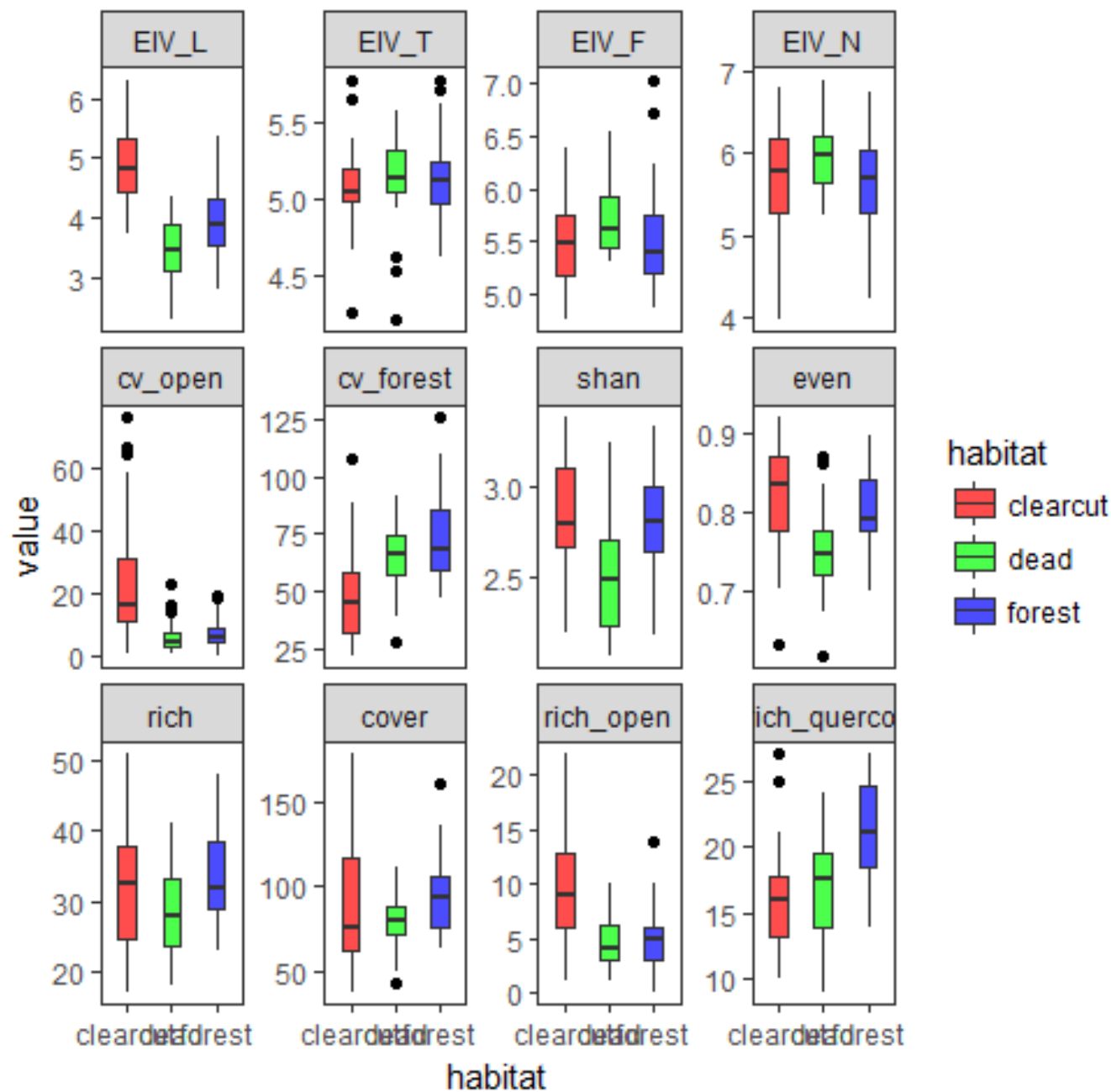
#przekształcenie do wąskiej tabelki

```
traity.hab.melt<-melt(traity.hab, id="habitat")
```

	habitat	variable	value
1	clearcut	EIV_L	5.2631579
2	clearcut	EIV_L	6.0000000
3	clearcut	EIV_L	4.9222222
4	clearcut	EIV_L	4.9175258
5	clearcut	EIV_L	4.8333333
6	clearcut	EIV_L	6.3125000
7	clearcut	EIV_L	5.7037037
8	clearcut	EIV_L	4.6153846
9	clearcut	EIV_L	4.6052632
10	clearcut	EIV_L	3.9491525
11	clearcut	EIV_L	4.8194444
12	clearcut	EIV_L	5.2600000
13	clearcut	EIV_L	4.4218750
14	clearcut	EIV_L	3.9687500

#tworzenie wykresów

```
kilka.wykresow<-ggplot(data = traity.hab.melt, aes(x =  
habitat, y = value, fill = habitat))+  
facet_wrap(~traity.hab.melt$variable, scales="free_y",  
nrow=3, ncol=4) +  
geom_boxplot(data=traity.hab.melt, mapping=aes(x=habitat,  
y=value),position=dodge,width=0.4, outlier.color="black")+  
theme_bw()+theme(panel.grid = element_blank())
```



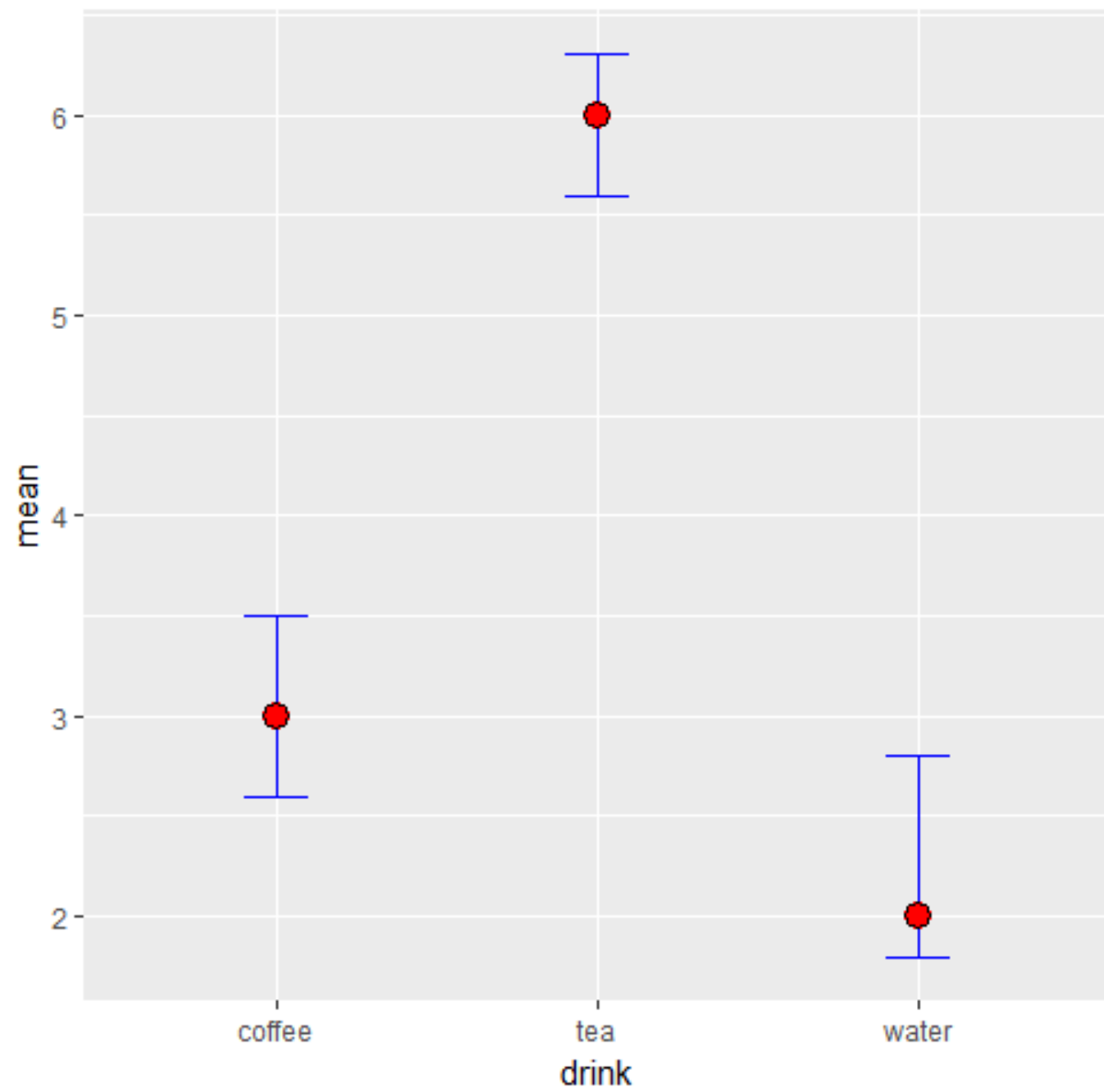
Wykres błędu

#tworzenie ramki danych

```
drinking=data.frame(drink=c("coffee","tea","water"),  
mean=c(3,6,2), lower=c(2.6,5.6,1.8), upper=c(3.5,6.3,2.8))
```

#tworzenie wykresu

```
ggplot() +  
  geom_errorbar(data=d, mapping=aes(x=drink, ymin=upper,  
ymax=lower), width=0.2, size=0.5, color="blue") +  
  geom_point(data=d, mapping=aes(x=drink, y=mean), size=4,  
shape=21, fill="red"))
```

Wykres słupkowy

#załadowanie danych do R

```
bar.cumul<-read.table('bar.cumul.csv', sep=';', dec=',',  
header=TRUE)
```

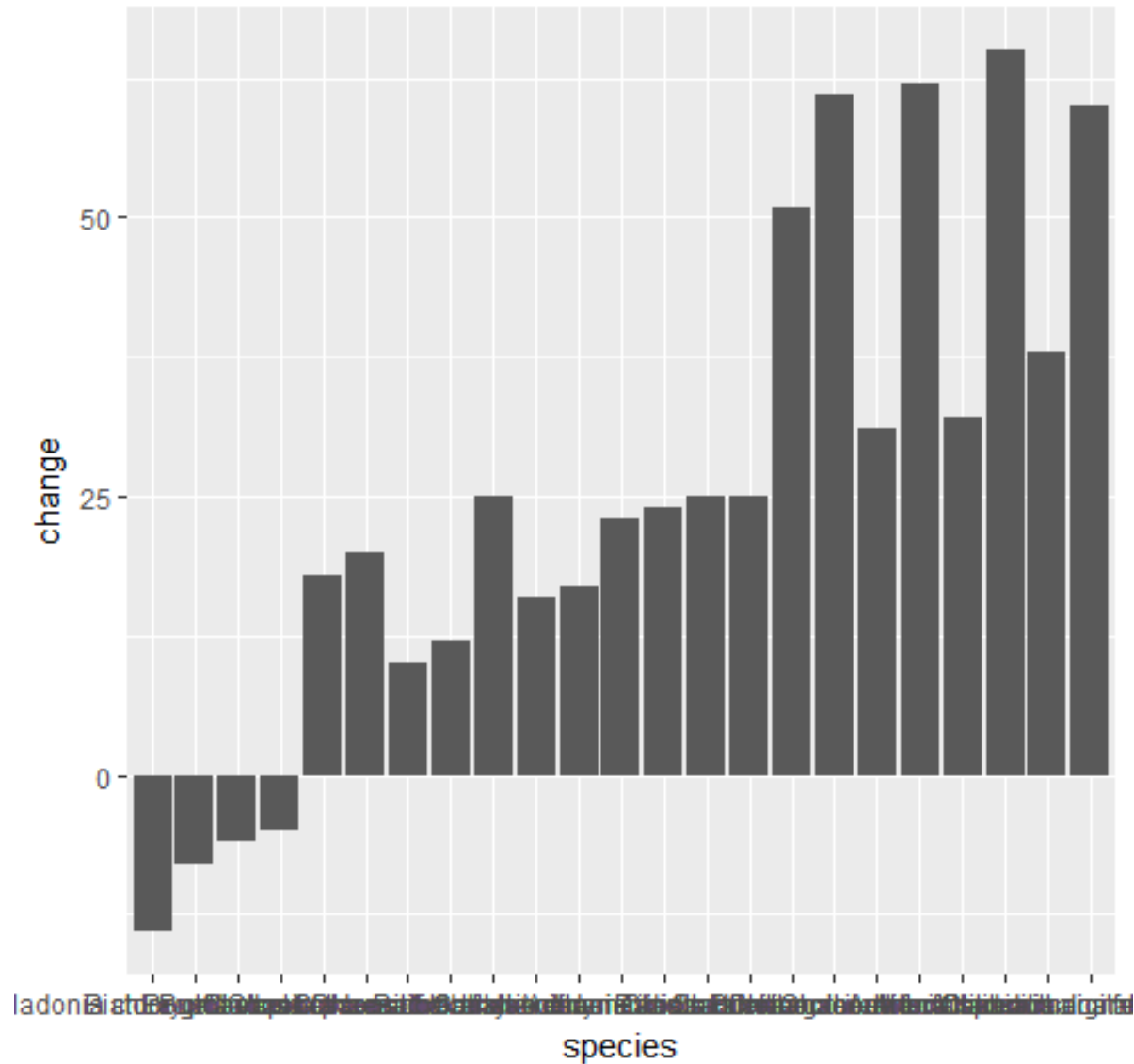
	species	forest	change
1	Anisomeridium biforme	decid	24
2	Arthonia didyma	decid	16
3	Arthonia spadicea	decid	32
4	Arthonia vinosa	decid	23
5	Arthothelium ruanum	conif	14
6	Arthothelium ruanum	decid	51
7	Bacidia laurocerasi	conif	7
8	Bacidia laurocerasi	decid	18
9	Bacidia rubella	decid	25
10	Biatora globulosa	decid	-8
11	Bryoria implexa	conif	-6
12	Buellia erubescens	decid	-5
13	Buellia griseovirens	conif	20
14	Buellia griseovirens	decid	41
15	Cetrelia cetrarioides s.l.	decid	17
16	Chaenotheca chlorella	conif	6
17	Chaenotheca chlorella	decid	14
18	Chaenotheca ferruginea	decid	60
19	Chaenotheca furfuracea	decid	31
20	Chaenotheca trichialis	conif	16
21	Chaenotheca trichialis	decid	46

#uszeregowanie gatunków względem wartości od największej do
najmniejszej (niewidoczne w ramce danych, ale widoczne na wykresie)

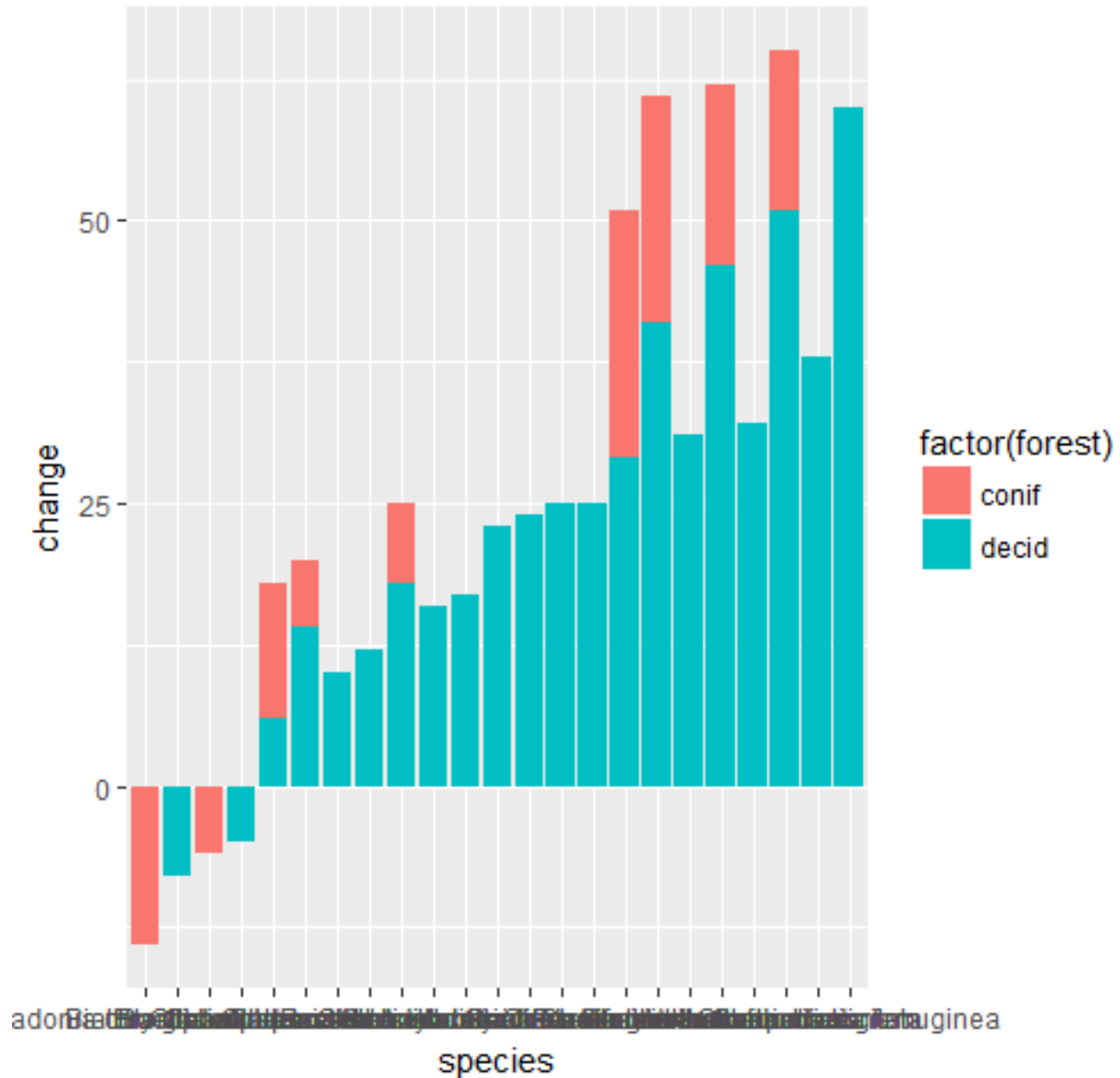
```
bar.cumul <- transform(bar.cumul, species = reorder(species,  
change))
```

```
#tworzenie wykresu
```

```
bar.plot<-ggplot(bar.cumul, aes(x=species, y=change))+  
  geom_bar(stat = "identity")
```

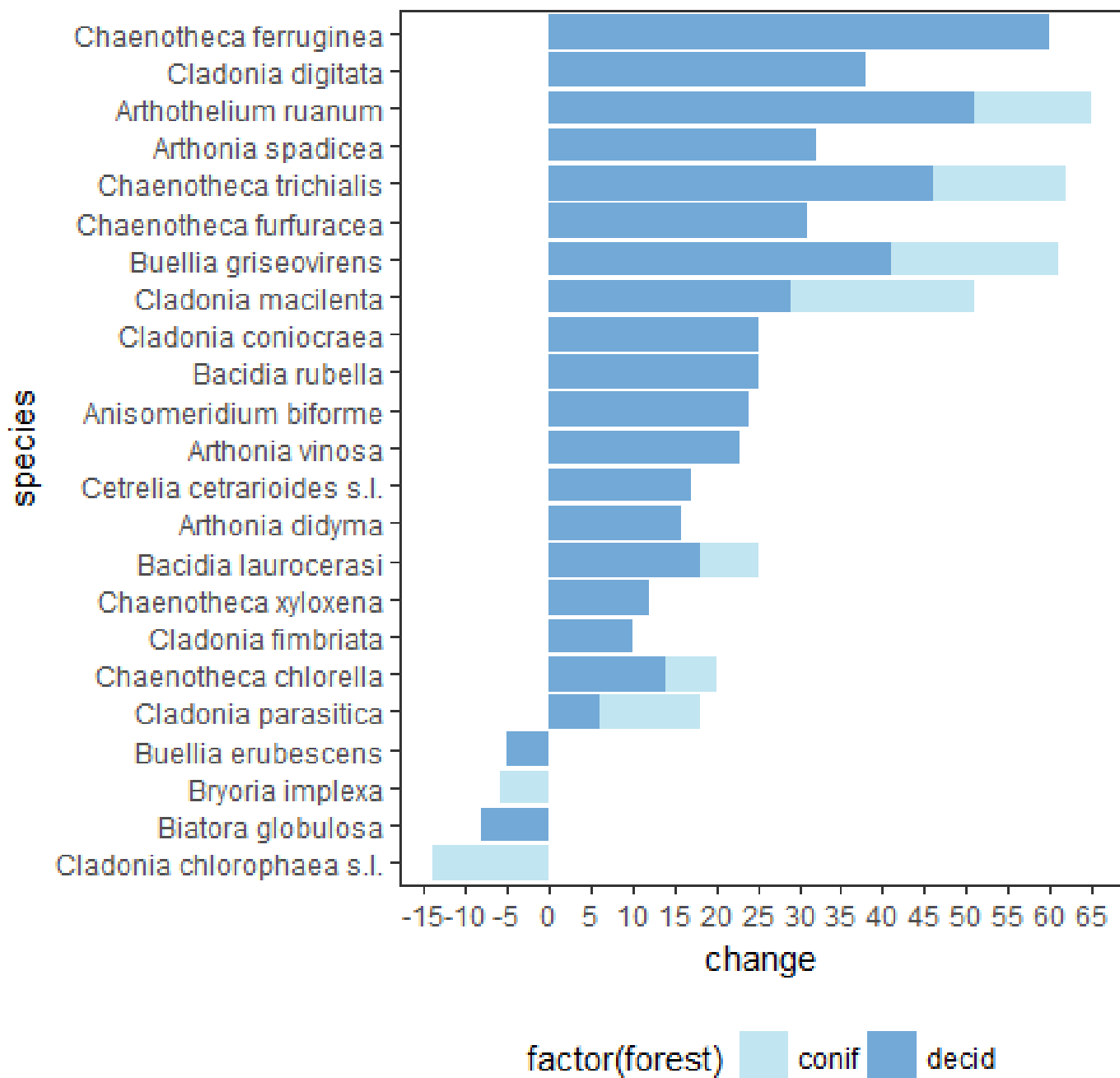


```
bar.plot<- ggplot(bar.cumul, aes(x=species, y=change,
fill= factor(forest)))+ geom_bar(stat = "identity")
```



#edycja wykresu

```
ggplot(bar.cumul, aes(x=species, y=change, fill=
factor(forest)))+
  geom_bar(stat = "identity") + coord_flip() +
  scale_fill_manual(values=alpha(c('lightblue2',
'steelblue3'), 0.8))+
  theme_bw()+
  theme(panel.grid = element_blank())+
  theme(legend.position = "bottom")+
  scale_y_continuous(breaks = seq(-50, 100, by = 5))
```



```
#wykres polarny (polarplot)
```

```
bar.plot+coord_polar()
```

