704SM Biostatistica

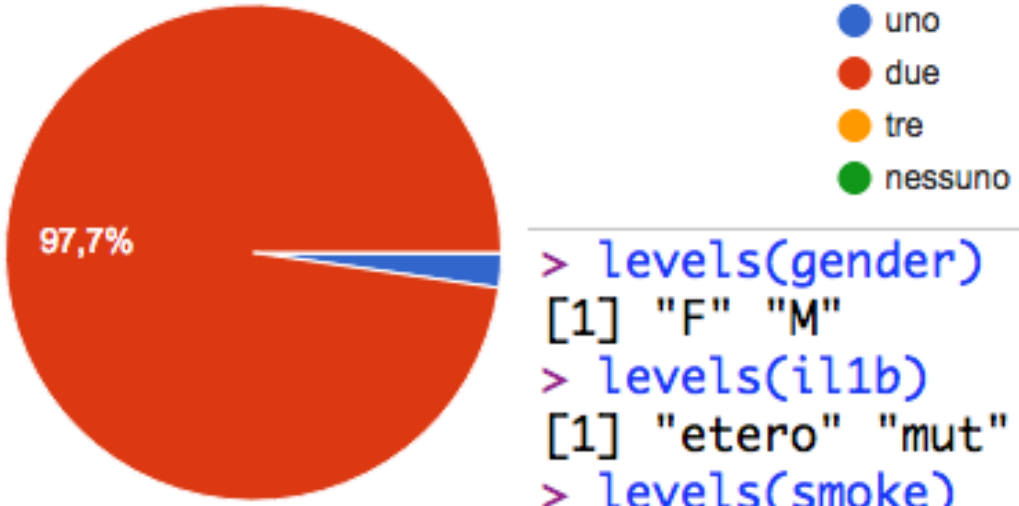
Massimo Borelli

ottobre 2016

UNIVERSITÀ DEGLI STUDI DI TRIESTE
Dipartimento di Scienze della Vita

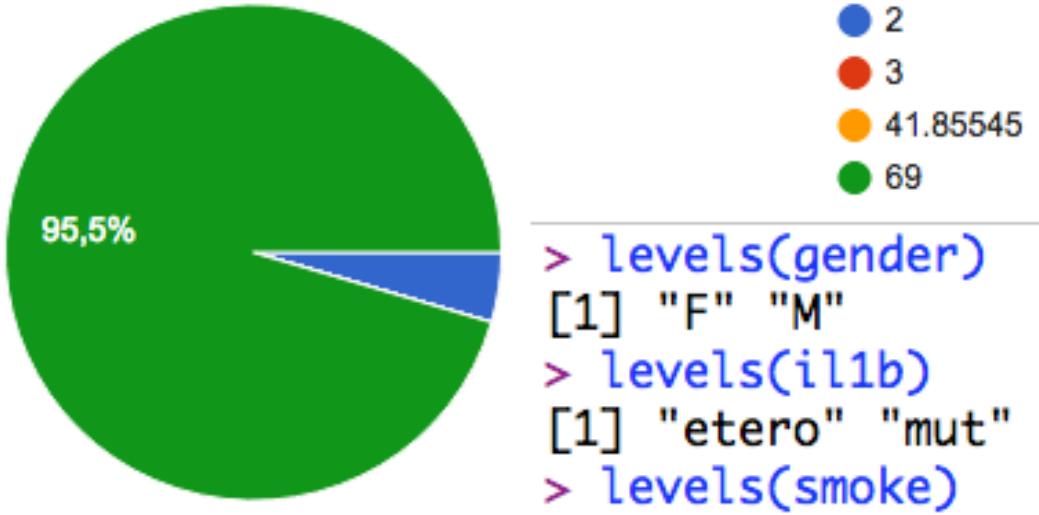
SOCIETA' DEI MATEMATICI
E NATURALISTI DI MODENA
www.socnatmatmo.unimore.it

Domanda 2: quanti livelli possiede la variabile gender? (44 risposte)



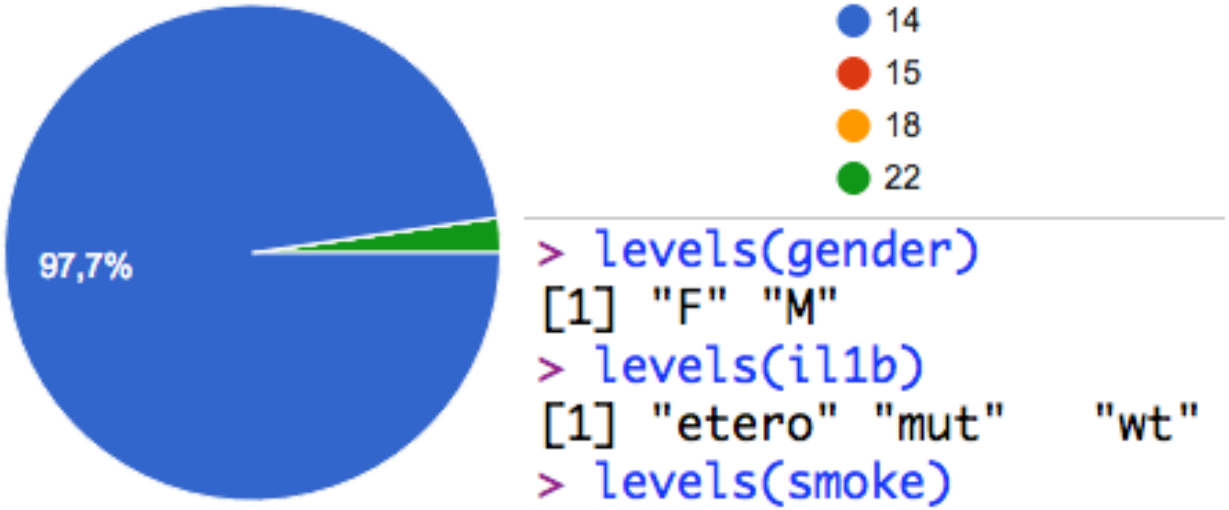

```
> levels(gender)
[1] "F" "M"
> levels(il1b)
[1] "etero" "mut"    "wt"
> levels(smoke)
[1] "high" "low"
> length(areainfl)
[1] 69
> mean(areainfl)
[1] 41.85545
> table(gender, smoke)
       smoke
gender high low
     F   14  22
     M   18  15
```

Domanda 5: quanti elementi possiede la variabile areainfl?



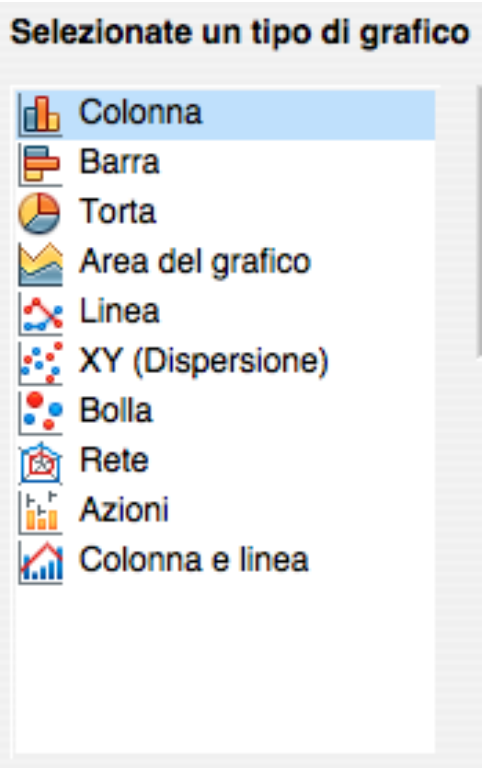

```
> levels(gender)
[1] "F" "M"
> levels(il1b)
[1] "etero" "mut"    "wt"
> levels(smoke)
[1] "high" "low"
> length(areainfl)
[1] 69
> mean(areainfl)
[1] 41.85545
> table(gender, smoke)
       smoke
gender high low
     F   14  22
     M   18  15
```

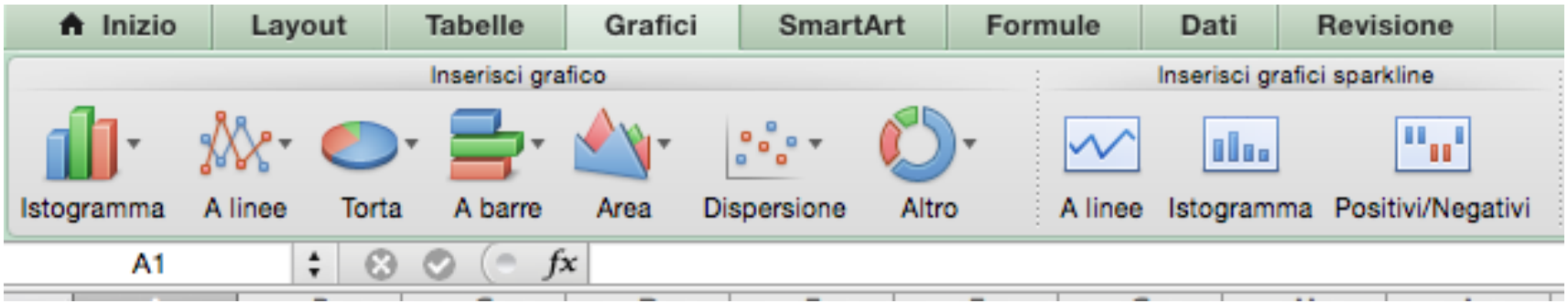
Domanda 7: quante sono le donne forti fumatrici? (44 risposte)




```
> levels(gender)
[1] "F" "M"
> levels(il1b)
[1] "etero" "mut"    "wt"
> levels(smoke)
[1] "high" "low"
> length(areainfl)
[1] 69
> mean(areainfl)
[1] 41.85545
> table(gender, smoke)
       smoke
gender high low
     F   14  22
     M   18  15
```

epic fail: diagramma a barre, non istogramma!

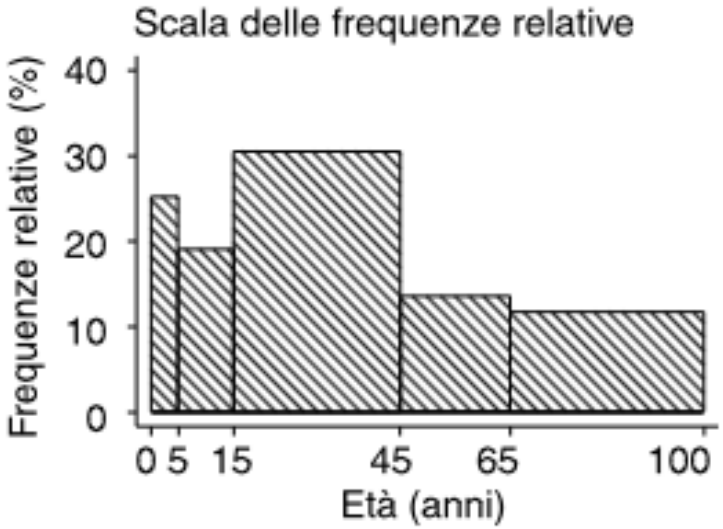


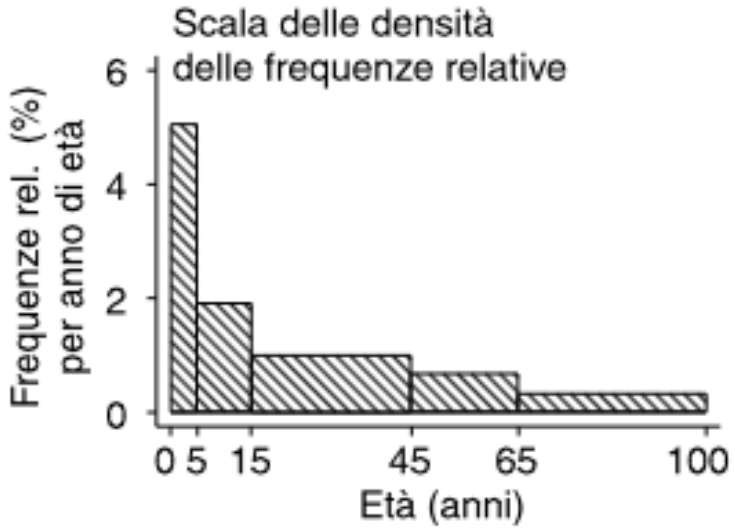


spieghiamo bene il concetto di istogramma

Tabella 4.7: Distribuzione dell'età relativa a persone vittime di incidenti domestici (Whittington 1977)

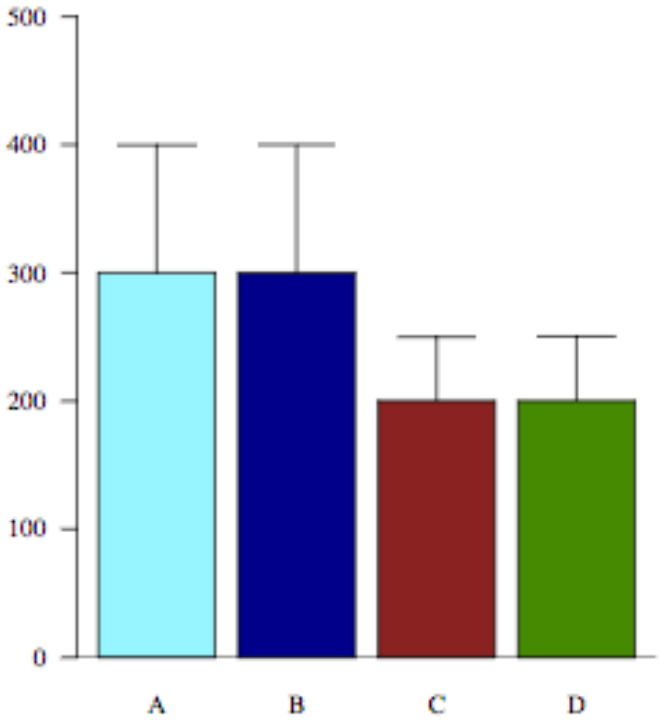
Gruppi di età	Frequenza relativa (%)	Frequenza relativa per anno (%)
0–4	25.3	5.06
5–14	18.9	1.89
15–44	30.3	1.01
45–64	13.6	0.68
65+	11.7	0.33





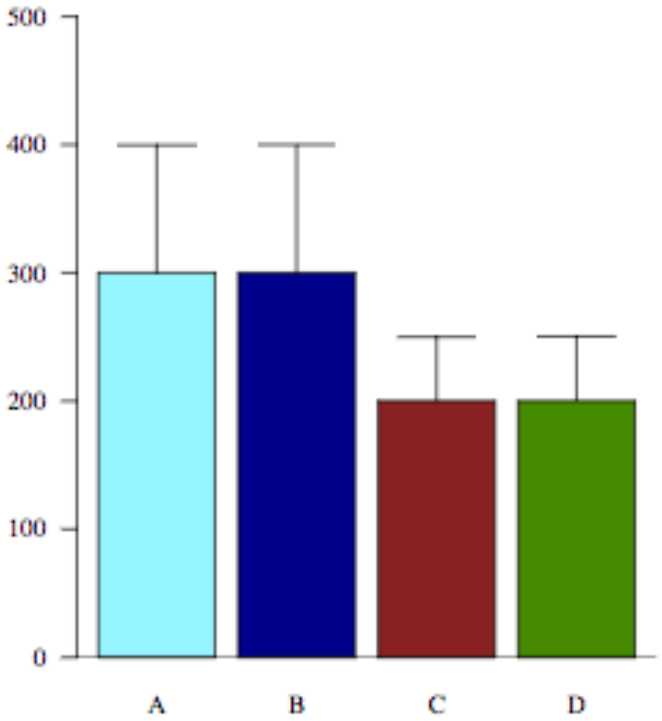
Why Are Dynamite Plots Bad?

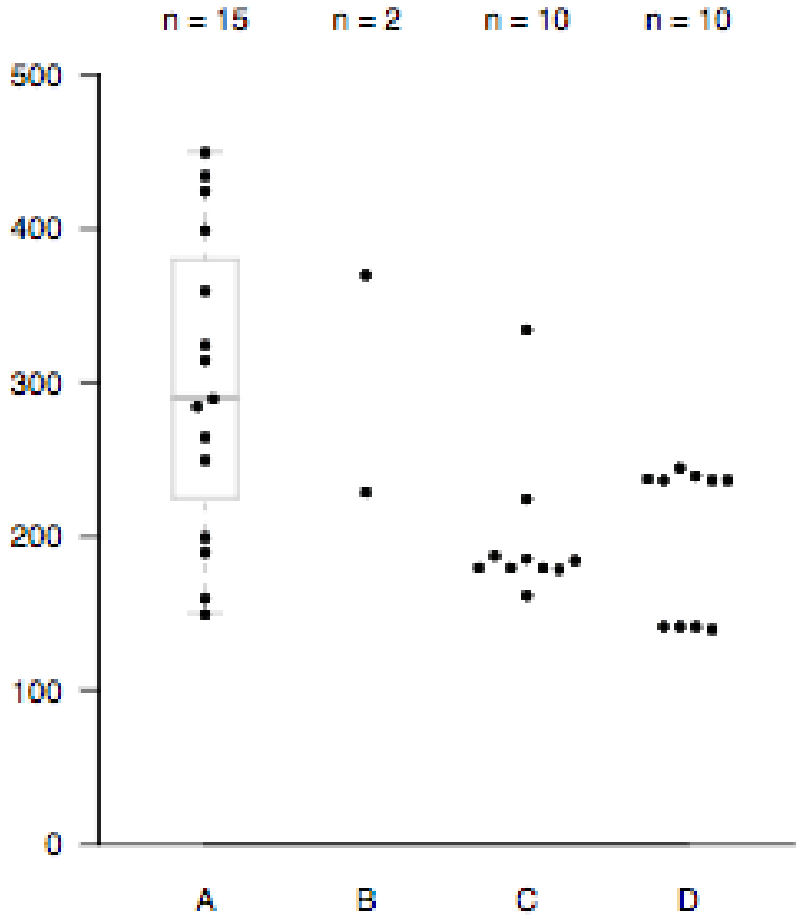
The height of the bar represents the average, in cartoons, don't they? Why are they bad?



Why Are Dynamite Plots Bad?

The height of the bar represents the average, in cartoons, don't they? Why are they bad?





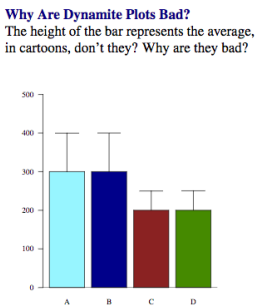
Tatsuki Koyama, 'Beware of Dynamite'

So little information! This plot presents 4 means and 4 standard deviations (or standard errors). This is a very inefficient use of space.

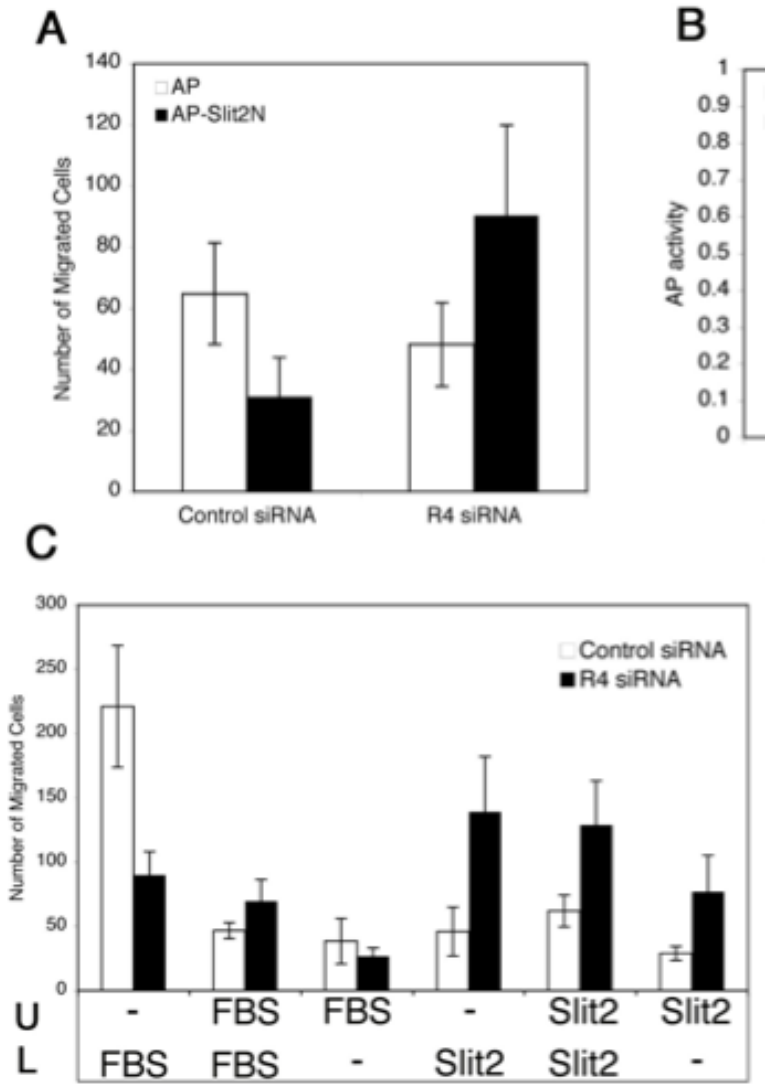
What do means mean? Averages do not usually convey much information. How spread are the data? Are there outliers? What are the sample sizes? None of these interesting and important questions is answered.

Whiskers get in the way. Whiskers add some information, maybe, but I don't know how to use them. They make the bars look taller, and the little information given by the bars is distorted.

Where are the data? There are data behind the bar, but the bar also covers space where no data exist. In addition, there must be some data above the bar in all that blank space.
Where are the data? You learn very little about the data and their *distribution*.



standard deviation o standard error?

Error bars in A (n = 3), and B (n = 3) represent SD while in C represent SEM (n = 4).

BMC Cell Biol. 2008 Nov 3;9:61.