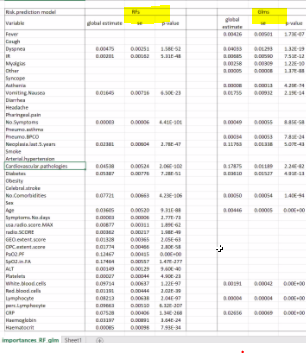
Slide 2

Ci sono tanti grafi al mondo e ognuno di loro ha delle particolari peculiarità.

Tutti questi chart potrebbero essere usati sullo stesso tipo di dati per far emergere cose differenti.

Slide 3

Quindi se ci sono i grafi, le tabelle non servono? NO! Anche le tabelle sono molto utili.

Slide 5

Quando si usa la tabella? Quando abbiamo tanti dati e siamo attenti al dettaglio.

Quando vogliamo vedere un particolare valore per un dato elemento.

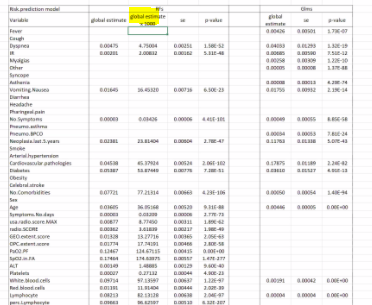
Si può voler vedere anche il rapporto tra le diverse variabili.

Viste in forma tabellare, in effetti può risultare difficile capire quale variabile è più rilevante delle altre.

Visivamente si può fare il confronto tra i due modelli (RFs e Gims evidenziati) e vedere ad occhio quali sono le variabili più rilevanti per un dato modello.

Non è tanto utile questa tabella, ma può risultare come ***look-up table*** (per i medici ad esempio).

[ **LOOKUP TABLE** – Tabella di ricerca per una semplice operazione di consultazione. ]



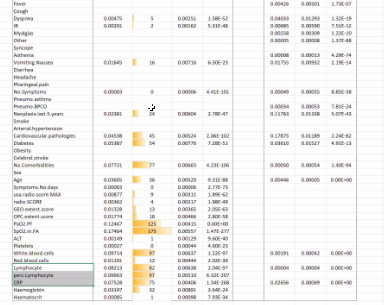
Come rendere questa tabella più utile?

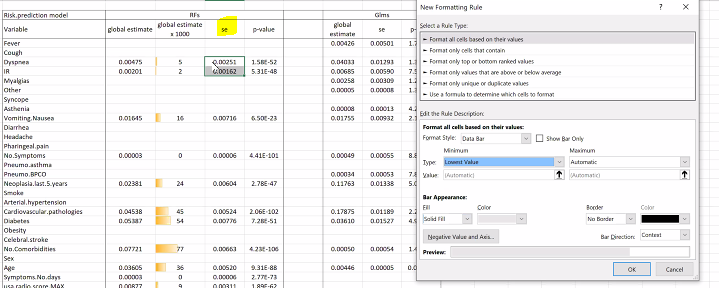
Intuitivamente si possono moltiplicare tutti i numeri (con 0,00…) magari per 1000 e si portano ad un’unità che visivamente piace di più:

indicare un titolo che dica cos’è stato fatto, ossia la molitplicazione per 1000.

Bene, visivamente i numeri sono più chiari.

Con Excel, selezionando la colonna evidenziata si può fare un **Data Bars** molto utile visivamente per capire la rilevanza dei vari valori.

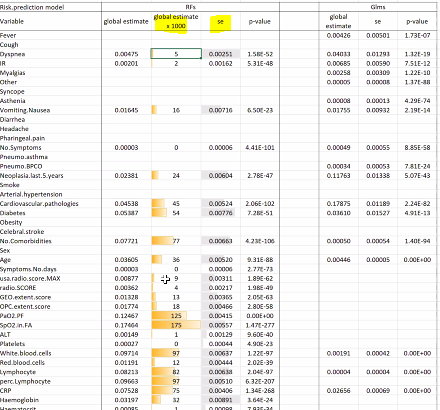
Questa visione fa capire bene cosa sta accadendo nella tabella.



Lo ***standard error*** (**SE**) indica quanto la stima gialla può variare.

Solitamente lo SE è meno importante e quindi lo si può visualizzare con un colore come il grigio.

Una visione così dice di non fidarsi molto dell’importanza → quindi la stima (gialla) non è molto rilevante.



Più è elevato lo SE più la stima (gialla) varia nel tempo.

Si tratta dello SE nella stima (l’**errore medio**): quanto varia la sitma intorno alla sua media.

Si utilizzano due colori diversi per indicare due concetti diversi e permettono di avere una visione d’insieme.

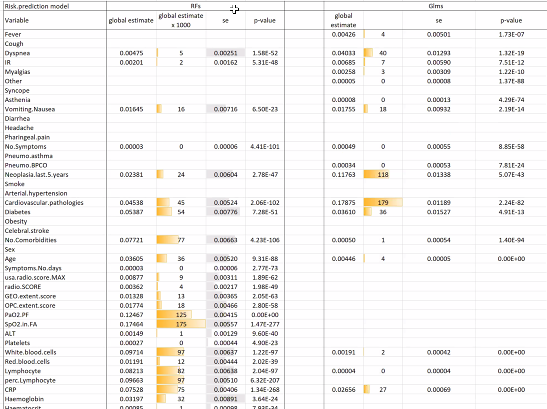
In questi dati è stata aggiunta una ***randomness*** per alcuni valori; quindi quando si fa una visualizzazione dei dati bisogna essere sicuri che i dati utilizzati siano ottenuti in modo consono, non in modo sfalsato altrimenti le visualizzazioni anche risultano sfalsate.

I dati devono raccontare il vero, dopo essere certi di questo, allora si visualizza.

Uno SE maggiore indica che c’è stata una variazione maggiore.

Si possono anche rielaborare i dati nel realizzare la visualizzazione; si può quindi (come si è fatto) aggiungere una colonna in cui si moltiplica per 1000.

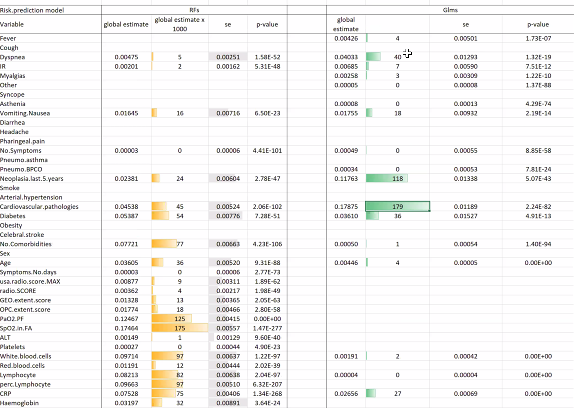
Si ripetono le stesse operazioni che si erano fatte a SX e si applica la Data Bar, esattamente come prima:



Si **comparano** così le due colonne gialle.

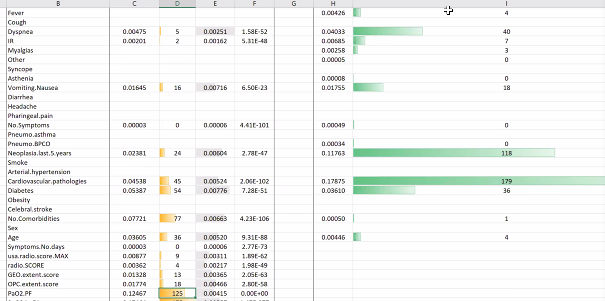
Essendo le due colonne però prodotte da due metodi diversi, si cambia il colore a DX.

Altrimenti l’utente che guarda la visualizzazione potrebbe pensare che il confrono avviene tra elementi corrispondenti, quando in realtà sono diverse:



Anche la larghezza della cella è fondamentale.

Vedere la cella più grande dà la percezione di importanza maggiore di quel dato:



Quando in realtà magari sia colonna gialla che verde hanno la medesima importanza.

Quindi si possono fare delle visualizzazioni all’interno delle tabelle con delle Data Bars ad esempio.

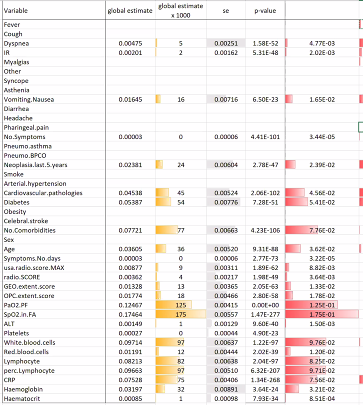
Per fare in modo che la tabella consenta ala comparazione tra un metodo ed un altro:

si inserisce una tabella e si può fare in modo di valutare l’importanza relativa di una variabile rispetto alle altre, ossia si fa in modo che la somma delle importanze sia pari a 1.

Si sommano tutti i valori della colonna gialla (le varie importanze metodo 1).

Si sta normalizzando ogni singolo valore. SI può dividere ogni valore per il totale.

**Normalizzando** tutti i valori, i data bars che si creano sono poi i seguenti:



Anche per l’altro metodo si applica la medesima normalizzazione: e tutti i valori della colonna verde (le varie importanze metodo 2) cambiano.

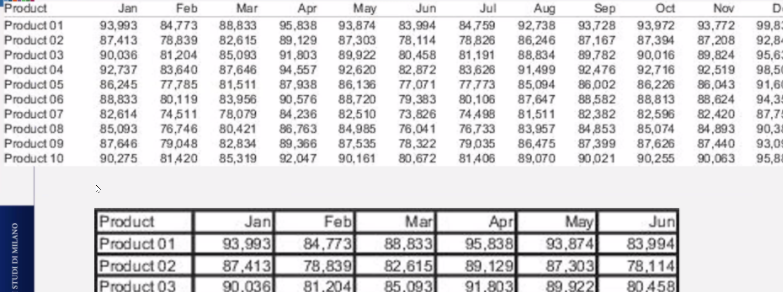
I data bars così ottenuti provengono da valori normalizzati:



Anche i data bars possono essere normalizzati, in modo da avere una visualizzazione vera, mostra il dettaglio nel particolare.

Mostra una visualizzazione che rende subito l’idea di ciò che intende comunicare, in modo chiaro.

Slide 6

Le tabelle devono avere un design che non sia molto pesante per chi le visualizza.

Ad esempio d’impatto la tabella sopra risulta più semplice da leggere.

Lo spazio è stato meglio gestito per fare in modo che i vari valori siano in modo migior confrontate.

Inoltre questo spazio impone al lettore di leggere per colonne (dato che l’uomo per abitudine legge per riga/di continuo).

(Tabella sotto)

Quando si trovano dei bordoni netti, si tende a concentrarsi su ogni singola cella.

E questo non va bene, perché la tabella è pensata per effettuare e vedere subito dei confronti.

In questa tabella sarebbe stato utile utilizzare un bold per indicare valori diversi o dei data bars per fare vedere il trend almeno.

Slide 7

È stato fatto in modo di ottenere barre dello stesso colore.

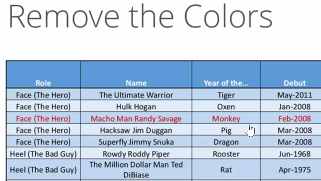
Il rosso indica il negativo, ossia la variabile è inversamente correlata alla presenza del Covid.

Questa visualizzazione ha permesso di dire che il metodo a DX, identificante dei predittori per prevedere se un paziente ha il Covid, non è attendibile.

Tramite impatto visivo si è capito di utilizzare le ***Random Forest*** (**RF**) come metodo predittivo.

Slide 8

**TABLE DESIGN** → rimuovere per migliorare

Qui si può vedere che è stata messa in risalto in rosso la riga rilevante.

Si tolgono allora i colori, tutto in scala di grigi.

Si rimuovono le linee.

Si rimuovono anche i colori alternati, tanto si può usare il bold per identificare il titolo delle celle.

Si potrebbe anche non preoccuparsi dei bordi, ma ci sono delle riviste che richiedono di utilizzare il loro stile.

In ogni caso, si tolgono i bordi.

Si rimuove il bold, tanto percettivamente si capisce che sopra il primo elemento è il titolo.

L’occidentale è abituato a leggere da SX a DX, ma così (come nella presentazione) risulta fastidioso leggere, impiega il 120% di tempo in più il nostro cervello a leggere così.

Quindi si allinea tuta a sinistra, non centrato.

Siamo abituati sin da piccoli a leggere in un certo modo quindi: tutti i testi allineati a SX e tutti i numeri allineati a DX.

Ci interessa di più che l’utente legga da SX a DX e confronti da SX a DX oppure ci interessa di più il confronto da DX a SX?

Se si vuole che il confronto avvenga tra i singoli gruppi → allora spostare i singoli gruppi: ciò impone percettivamente all’utente di confrontare in verticale.

Lo spazio tra i gruppi fa in modo che si faccia il confronto tra gruppi.

In alcuni punti ci sono 3 cifre decimale, in altri una cifra decimale perciò si utilizza un formato consistente in modo che piacca all’utente, soprattutto visivamente: si utilizza una sola cifra decimale, si moltiplica e si rendono tute le cifre allo stesso formato.

Una volta identificato i gruppi e imposto all’utente di guardare poi ogni singolo gruppo, si rimuovono prima le ripetizioni.

In tutto, anche nelle visualizzazioni, il font del testo è fondamentale.

Ci sono caratteri ***serif*** e ***sans serif***:

• *sans serif*: sono i caratteri con le grazie (vedi Calibri)

• *serif*: sono i caratteri senza le grazie (vedi Times New Roman)

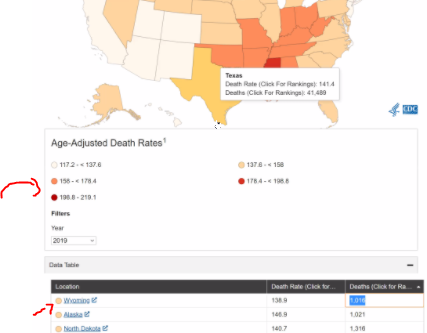
Calibri, Roboto sono *sans serif* più semplici da leggere (anche per le stampanti sono migliori, consumano meno inchiostro).

Questa scelta del font invoglia il lettore a leggere la tabella.

Quindi secondo il principio di Tufte: il meno è più attrattivo ed è più di impatto.

Slide 9

Va bene avere delle tabelle ma se non è necessario analizzare dei valori nel dettaglio, si possono utilizzare anche altri strumenti.

Slide 10

In quest mappa, non solo si ha una tabella che riporta il *mortality rate*, ma anche il numero di morti.

Il colore accanto alla nazione (con tanto di legenda per spiegare → TUTTO DEVE ESSERE MASSIMAMENTE SPIEGATO) è informativo, indica il raapporto di morti per ogni nazione rispetto alla popolazione della singola nazione.

Prima di plottare i numeri bisogna essere sicuri che non siano non veri o sfalsati.

Non si può fare la conta assoluta del numero di morti per ogni nazione, ma è necessario normalizzare (per il totale della popolazione).

Solitamente ormai si normalizza rispetto a 100 000, indicante la popolazione totale.

La normalizzazione si può fare rispetto a diverse cose.

Quindi la mappa è accompagnata da una tabella.

Slide 11

Nello stesso momento si sta visualizzando: la dimensione della provincia, la sua forma, quindi per ogni provincia si stanno vedendo 4 dimensioni (forma, dimensione, *cancer rate*…).

Se dovessi guardare una tabella che porta 3056 valori, ora che riesco a identificare come zona di alto pericolo per i vari tumori delle date province, si fa fatica.

Nello stesso tempo queste tre tabelle dicono che c’è un dato afflusso di un dato tumore in una certa zona della mappa.

Per dire in quella zona ci sono tanti malati di tracheite, allora si vede magari se ci sono delle particolari aziende.

Slide 14

**Normalizzazioe Edge adjusted**

SI va a vedere la popolazione e si normalizza in termini di percentuale di presenza.

Slide 15

Si misurano gli eventi che succedono per i bambini dagli 1 ai 4 anni, inoltre ogni gruppetto (1-4, 5-9..) può indicare la specifica provincia quindi, per dire, 1-4 può indicare che i bamibini dagli 1 a 4 anni nel Texas (1° posizione del gruppetto 1-4 nella lista) sono affetti da una data malattia.

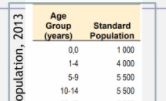
Slide 17

Si prende il numero di persone in quell’aera geografica e la si moltiplica per 100 000.

Supponendo che quell’area geografica sia abitata da 100 000 persone.

Slide 19

Si va a vedere di questi 100 000 abitanti, quanti in percentuale hanno 0 anni, quanti 1,….

La *standard population* si ottiene dividendo:

1-4 → 1 diviso 100 000

4 diviso 100 000

La somma di tutte le *standard population* risulta 1.

Slide 23

Il 70% degli stimoli importanti provengono/vengono acquisiti dall’occhio.

Il resto viene distribuito dagli altri sensi.

Basti pensare al fatto che quando mangiamo, il colore del cibo ha la sua valenza:

per dire il risotto giallo piace tendenzialmente a tutti perché il giallo è un colore che indica calore, il sole…cose belle insomma.