

5/3.

a) Athletes – Bar Chart.

- Töltsd le az atlheltes.xlsx-et a Teams fájlok fölről vagy a holnapomról, majd töltsd be Tableau-ba.
- Konvertáld át azokat a típusokat ahol van értelme, például a mennyiségi változókat. Ezt úgy kell, hogy még a Data Source felületen a változó fejlécénél az Abc ikonra rá kell kattintani és a megfelelő típust ki kell választani.
- New Sheet. Adatoknál baloldalon, ami szám azt konvertáld át Measure-re.
A változó job oldalán kisnyíl -> Convert to Measure.
- Hozz létre egy új szűrőt nemzet szerint. Alapból egyik nemzet se legyen kiválasztva. Aztán Show Filter és a kisnyíl menünél válassz ki egy értelmes szűrés típust, például Single Value Dropdown. Itt lehet gépelve keresni is, válaszd ki Magyarországot.
- Hozz létre egy Calculated Fieldet, ami legyen TRUE ha valaki többszörös érmes illetve legyen FALSE különben
(Segítség: Csak azt kell ideírni, amit egy if állítás feltételébe írnál, de használhatod a Tableau IF függvényét is, ez esetben a TRUE/FALSE értékek helyett lehet bármi más bináris.)
Eszerint az új mező szerint szűrd ki azokat a sportolókat, akik többszörös érmesek.
- Végül készíts egy oszlopdiagramot, az oszlopoknál legyenek a nevek, a soroknál pedig a három érem típus.
Színezd be a különböző éremtípusokat különböző színűekre. Mondjuk kb arany ezüst és bronz színűre.

b) Scatter plot.

- Hozz létre egy új munkalapot. Az egyik tengelyen legyen a weight másikon pedig a height és minden pont feleljen meg egy sportolónak.
- Szűrd ki a következő 3 sportot: Basketball, Cycling, Weightlifting illetve szűrj nem szerint is, tekintsük csak a férfiakat.
- Színezd az ábrát a Sport szerint majd állítsd át a pontok alakzatát teli karikákra (shape) és a színeknél állítsd be, hogy átlátszóak legyenek a pontok (opacity = 60%)
- Állítsd át a tengelyek tartományát (tengelyeken jobb gomb, edit). A weight tengely tartománya kezdődjön 50-től, a height pedig 1,42 től.
- Látható, hogy a jobb alsó sarokban ott van hogy 19 Null. Erre rákattintva szűrd ki a null értékeket.
- Határozd meg, hogy hány évesek voltak az olimpia idején. Ehhez használd a date és a datediff függvényeket. Ahhoz hogy le tudd tesztelni, hogy működik-e, hozz létre egy új munkalapot és a sorokra a neveket tedd, majd az ábrán a nevek mellé húzd az Abc-k helyére a kiszámolt korokat, vagy ugyanezt az eredményt kapod, ha Marks kártyán belüli Text-re teszed rá a mezőt.
- Az előbbi szórásdiagramon húzd rá a kiszámolt életkort a Tooltip-re.

A tooltip arra jó, hogy további információkat adjunk az ábrához (az újabb verziókban már nem csak szöveget, hanem az adott pontra vonatkozó vizualizációt is lehet tooltip-be tenni).

- A Tooltip-re rákattintva be lehet állítani a megjelenést. Állítsd át olyanra, hogy amikor ráhúzzátok az egeret egy pontra az jelenjen meg hogy "sportoló neve" competed in "sport neve" at the age of "életkora az olimpián". Illetve az átkötő szöveg ne legyen félkövér, de a lényeges adatok pedig legyenek narancssárgák és mondjuk 12 es betűméretűek.

Azaz egy konkrét sportolónál: **Daniel Martin** competed in **cycling** at the age of **22**

(Megjegyzés: Egy jól megszerkesztett Tooltip-ről az embereknek az a benyomásuk, hogy aki ennyi időt töltött egy ilyen jelentéktelen dolog megszerkesztésével, az biztosan az elemzésekbe is sok időt fektetett)

c) Paraméterek.

- A Sample - Superstore adatokat használva hozz létre egy vonaldiagramot, ami a heti eladott mennyiségeket mutatja (Quantity).
- Most szeretnénk az ábrához adni egy referencia vonalat, hogy lássuk melyik hetek hogyan teljesítettek a referencia vonalhoz képest. Továbbá azt is szeretnénk hogy egy külső paraméterrel lehessen állítani a referencia vonal magasságát.
Ehhez először hozz létre egy paramétert: A Dimensions felirat melletti kisnyíl, Create Parameter. A Data Type legyen integer, mivel az eladott mennyiség is egész szám, a megengedett értékeknek pedig adj meg egy értelmes tartományt. Az ábra alapján ez legyen mondjuk 0 és 550 között. Ha ez kész a Measure alatt megjelent ez az új paraméter. A parameter job oldalán kisnyíl: Show Parameter Control, vagy az Analysis menün belül a Parameters menüpontban is meg lehet jeleníteni. Ha a most megjelent csúszkán átállítod a parameter értékét, még nem történik semmi, mert nincs összekötve a paramétert semmivel.
- Hozz létre egy referencia vonalat (Analytics pane, Constant Line), ami bemenetként az előbb létrehozott paramétert használja. A referencia vonalon a paraméter értéke szerepeljen.
- Hozz létre egy Calculated Field-et, ami TRUE ha az összes eladott mennyiség sum(quantity) nagyobb, mint a referencia vonal értéke (azaz a parameter értéke). Majd eszerint az új mező szerint színezd az ábrát. Nézd meg mi történik most, ha a csúszkán átállítod a parameter értékét.

d) Histogram

- Hozz létre a magasság eloszlására egy hisztogramot. (A Show Me felületről ezt könnyen meg lehet tenni, csak ki kell jelölni a változót, majd a Show Me felületen kiválasztani a hisztogramot).
- Húzd a nemeket tartalmazó mezőt a színekre. Majd a Marks kártyán állítsd át az alakját Bar-ról Polygon-ra. Próbáld ki hogy a többi alakkal milyen ábrákat kapsz, pl Area, Line. A színeknél állítsd be hogy áttetsző legyen.

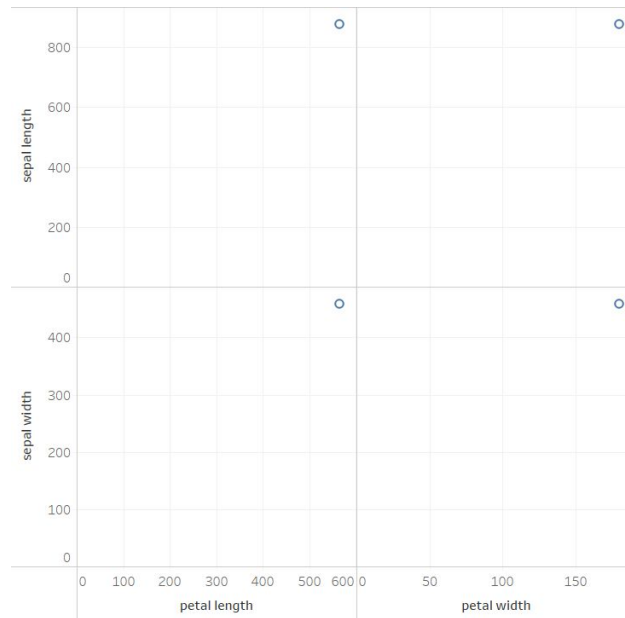
- Most azt szeretnénk, hogy egy külső állítható paraméterrel tudjuk állítani a histogramhoz használt bin-ek méretét. Ehhez hozz létre egy paramétert, a neve legyen mondjuk Bin Size, típusa float, és a megengedett értékek pedig legyenek egy tartományon belül (range) Min: 0,011 Max: 0,154 és a Step Size pedig szintén 0,011.
- Amikor létrehoztuk a histogramot, létrejött egy height (bin) nevű dimenzió (ha nem a Show Me felületet használod, akkor manuálisan kell létrehozni egy ilyen változót). Ezen jobbgomb/kisnyíl Edit: Size of binst nyissuk le és válasszuk ki az újonnan létrehozott Bin size paramétert.
Megjegyzés innen közvetlenül is létrehozhattuk volna a paramétert, a Create New Parameter... -re kattinva.
- Most jobboldalon állíthatjuk a binek méretét és ennek megfelelően változik az ábra. (Az alapbeállítás eredetileg 0,066 volt)

e) Linear regression.

- Töltsd be a Sample – Superstore fájlt, majd ábrázold vonaldiagramon a Profitot (a datum változó legyen az Order date). A dátumot állítsd át folytonos, negyedéves felbontásra.
- Add hozzá az ábrához a lineáris regressziót. (Ehhez először létre kell hozni egy Calculated Field-et. Segítség a honlapon a Tableau Calculated Fields-nél)
- Állítsd be duális tengelyeket (dual axis) és aztán szinkronizáld a tengelyeket.
- módosítsd a Lineáris Regressziót úgy, hogy a magyarázó változó csak a Sales legyen, és az aggregálást változtasd meg átlagról összegre.
- A lineáris regresszió vonalának vastagsága legyen az abszolút eltérés logaritmusának nagysága. (Ehhez létre kell hoznod egy új Calculated Field-et ami kiszámolja a Profit és a lineáris regresszió különbségének abszolútértékének a logaritmusát, és ezt a mezőt kell a Size kártyára helyezni)

f) Clusters.

- Töltsd le az iris.xls file-t. Ábrázold pontdiagramon a négy measure-t (mondjuk a sorban legyenek a petal length,width változók, az oszlopban pedig a sepal length width). Ha ez megvan akkor az ábrának valahogy így kell kinéznie:



Mivel a Tableau alapértelmezetten minden measure-t aggregál, ezért most a négy pont a measure mezők összegeit mutatja (az alapértelmezett aggregálás az összeg, de ezt át lehet állítani).

- Az Analysis menüpontban van egy olyan lehetőség hogy Aggregate Measures. Erre kattints rá és most már minden pont egy sornak felel meg az adattáblából.
(Megjegyzés: Az aggregálást általában nem érdemes kikapcsolgatni mert nem biztos hogy az történik, amit szeretnénk, jobb ha a pontfelhődiagramnál a detail-re helyezünk mezőket, de ebben az esetben az adatforrás jellege miatt ezt megtehetjük, másrészt nem is lehet semmit a detail-re beállítani amivel ugyanezt az ábrát kapnánk)
- Hozz létre egy Calculated Field-et ami az R-en keresztül klaszterezi a pontokat (kmeans)
(Segítség: az első kód a honlapon: <http://math.bme.hu/~marcessz/scripts.html>)
- A kiszámolt klaszterek szerint színezd az ábrát. Ha folytonos lett a színezés, akkor konvertáld át a kiszámolt mezőt diszkrétre.