

# Tableau

## Основне информације и инсталација

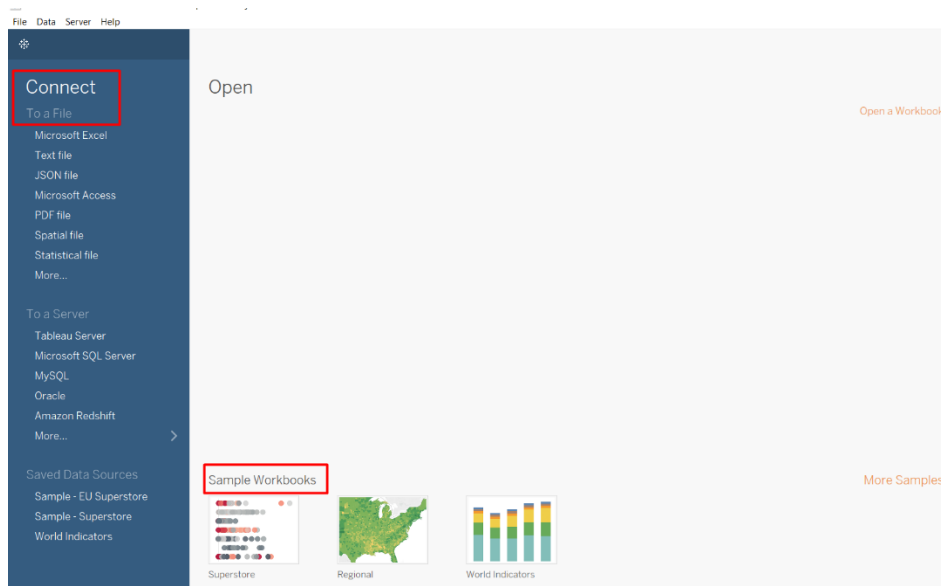
Tableau је софтвер намењен анализи и визуелизацији података и за те потребе се користи како у професионалне сврхе у компанијама, тако и за академска истраживања. У индустрији је намењен коришћењу од стране топ менаџмента за потребе пословног извештавања на ефикасан и једноставан начин.

На сајту (<https://www.tableau.com>) се могу пронаћи различити производи овог програма (*desktop* и *web* верзија; програм за саму припрему података и Tableau Server) као и доста корисних савета за његово коришћење. У даљем тексту фокус ће бити на раду у *desktop* верзији овог програма.

Приликом инсталирања треба да упишете личне податке као и податке везано за сврху коришћења и компанију/факултет у оквиру које ћете радити истраживање. Када је програм успешно инсталиран имаћете 14 дана бесплатног коришћења након чега ће вам требати *license key* за даље коришћење.

Сад је све спремно и можемо да се упознамо са интерфејсом и основним појмовима везано за Tableau.

## Упознавање са програмом



У оквиру главног прозора можете видети пример *Workbook-a*. Он представља структура пројекта овог софтвера, а *.twbx* је формат у коме га треба сачувати када је готов.

Са леве стране је мени за одабир и учитавање података над којима ће се радити анализа. По том питању пружају се многе могућности везано за формат података. То могу бити SQL табеле, Excel табеле, JSON и XML фајлови, csv подаци итд.

У овом примеру биће коришћен csv формат са портала отворених података, а везано за анализу квалитета воде у Србији 2016. године. Податке можете преузети овде (<https://data.gov.rs/sr/datasets/kvalitet-voda>).

Када смо учитали податке наш главни прозор изгледа другачије. Са леве стране се налазе табеле које су нам на располагању (тренутно је ту само једна, али може да их има више нпр.у случају када Excel фајл садржи више *sheet-ова* или SQL база има више од једне табеле што је обично и случај).

The screenshot shows the Tableau interface with a connection to a CSV file named 'kvalitet\_vode\_2016.csv'. The interface includes a menu bar (File, Data, Server, Window, Help), a toolbar, and a sidebar with 'Connections' and 'Files' sections. The main area displays a data table with columns for station ID, name, date, and various water quality parameters.

Sifra Stanice	Naziv Stanice	Datum Uzorkovanja	DubinaUzorkovanja...	Proticaj m3/s	Vodostaj cm	Vodostaj m.n.m	Temperatura vode...	Temperatura vazd...	VidljiveOtpadneM...	Minis opis	Boja opis
42,010	Bezdan	1/13/2016		50	1,210.00	-7	null	3.3000	4.0000	bez	bez
42,010	Bezdan	2/9/2016		50	1,950.00	129	null	4.3000	8.0000	bez	bez

Већина акција при коришћењу Tableau-а се заснива на *“drag and drop”* механизму, што у буквалном преводу значи *“превуци и остави”*. Тако да ће одабрана база бити изабрана кликом миша и довођењем у главни прозор. Да постоје још неке табеле истим принципом би и њих довели у главни мени стим што бисмо морали да изаберемо по ком атрибуту бисмо спојили те две табеле (то је углавном неки ID, а од предмета анализе зависи који ћемо конкретно да изаберемо). Ако сте се сусретали са релационим базама података овај механизам неће бити велики изазов.

Када су подаци импортовани у горњем десном углу треба да изаберете начин рада над њима. Опције које су понуђене су *Live* и *Extract*.

- **Live** рад представља такав рад да ако дође до неке промене у коришћеном фајлу аутоматски ће се променити и у активном *Workbook*-у
- **Extract** метод чини да подаци који ће се анализирати буду исти као у тренутку њиховог учитавања. У зависности од намене израде извештаја, количине података и личних преференција везано за време извршавања наредби треба да изаберете једну од ове две опције.

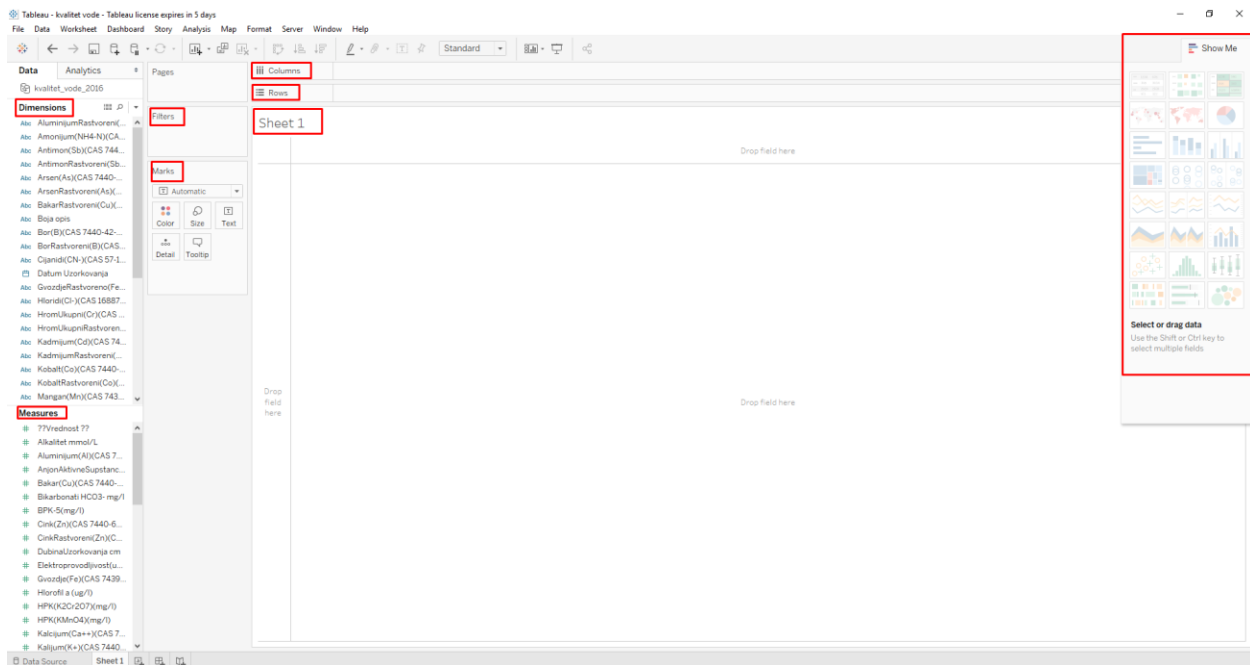
У овом случају ћемо користити Extract јер ће анализа бити једнократна и желимо да је учинимо што бржом.

Након овога пре свега се треба упознати са типовима података које имамо у бази. Генерална подела је на мере и димензије.

- **Мере** су у највећем броју случајева нумерички подаци (континуалне вредности) који могу да се агрегирају, тј. да се спроводи одређена операција над њима као што је рачунање суме, просека, минималне и максималне вредности итд. Као такве оне представљају главне показатеље и у извештајима се детаљније објашњавају кроз изабране димензије.
- С друге стране **димензије** су категоричке (дискретне) вредности. Оне су потребне да ближе опишу мере и могу бити: временске, просторне и остале димензије. Ако погледате свој *data source* (извор података) видећете да је сваком атрибуту додељена сличица као симбол којој врсти података припада.

Ако постоје атрибути за које смо сигурни да нам неће користити при анализи можемо их искључити кликом на тај атрибут и изабрати опцију *hide*. У нашем примеру искључили смо све бројевне податке које у својим вредностима садрже ознаку < или > зато што као такве нису више нумеричке већ категоричке. Такође, смо смањили обим података, што иначе не би требало да се ради јер истраживање губи на квалитету, али с обзиром да је циљ да се покажу функционалности овог програма, то је легитимно учинити јер су нам на располагању остали многи нумерички атрибути. Друга опција када наиђете на специјалне карактере у нумеричким атрибутима је такође на поље тог атрибута да кликнете на опцију *custom split* и изаберете симбол у односу на који желите да издвојите вредности (у овом случају <). На тај начин ћете креирати дупликат тог атрибута без специјалних карактера. Затим кликом на иконицу која означава тип атрибута (Abc) промените на *numerical, decimal*. Након овога можете обрисати оригиналну варијаблу и надаље користити правилно форматирану копију.

Такође ради каснијег коришћења географских одредница кликом на атрибут "Назив станице" и одабиром опције *Aliases* дефинисани су нови (реални) називи станица без специјалних карактера уместо слова *č, ć* и *ž*. Тако на пример село *Kusiae* добија *alias-a* *Kusic*.



Након што су подаци уčitани и сређени може се почети са обрадом и анализом истих.

Као што се може приметити на слици са леве стране се налазе већ поменуте мере и димензије разврстане по тим категоријама. Са десне стране се налазе понуђене визуелизације и преласком миша преко сваке ће се појавити колико је мера, а колико димензија потребно да би она била искоришћена.

Нови појмови за даљи рад су *sheet*, *dashbord* и *story*.

- **Sheet** представља оно што се види на претходној слици, дакле “простор” на коме се може направити једна визуелизација;
- **Dashbord** је “табла” која служи за упоредни приказ више *sheet* -а, а као посебна функционалност је та да на једном *dashbord*-у *sheet*-ови садрже бар једну заједничку димензију по којој се може филтрирати приказе (нпр. по територији или за одређени период).
- **Story** представља коначни приказ креираних визуелизација, како *sheet*-ова тако и *dashbord* -а где поред графика свака “страница” *story*-ја може садржати и текст у виду описа и закључака везано за анализу.

Уоквирене црвеном бојом су и ствари које ће се најчешће користити.

- **Columns** - представља варијаблу која ће бити приказана вертикално на нашем *sheet*-у,
- **Rows** - оне које ће бити приказане хоризонтално

И димензије и мере могу бити колоне и редови, али је због прегледности неписано правило да димензије буду рапоређене по редовима, а мера по колонама. Заједничким именом атрибути који се смештају у поља колона и редова се називају “pills” због облика који ће придобити.

- **Filter** је веома корисна функционалност када нека димензија узима доста вредности (нпр. када имамо 100 градова). Филтер нам омогућава да анализирамо одабрану меру по једном, неколико или свим градовима.
- Опција **Marks** нам омогућава бољи визуелни приказ мера и димензија (тако можемо температуру ближе приказати кроз боју – што је температура већа, то је боја тамнија).

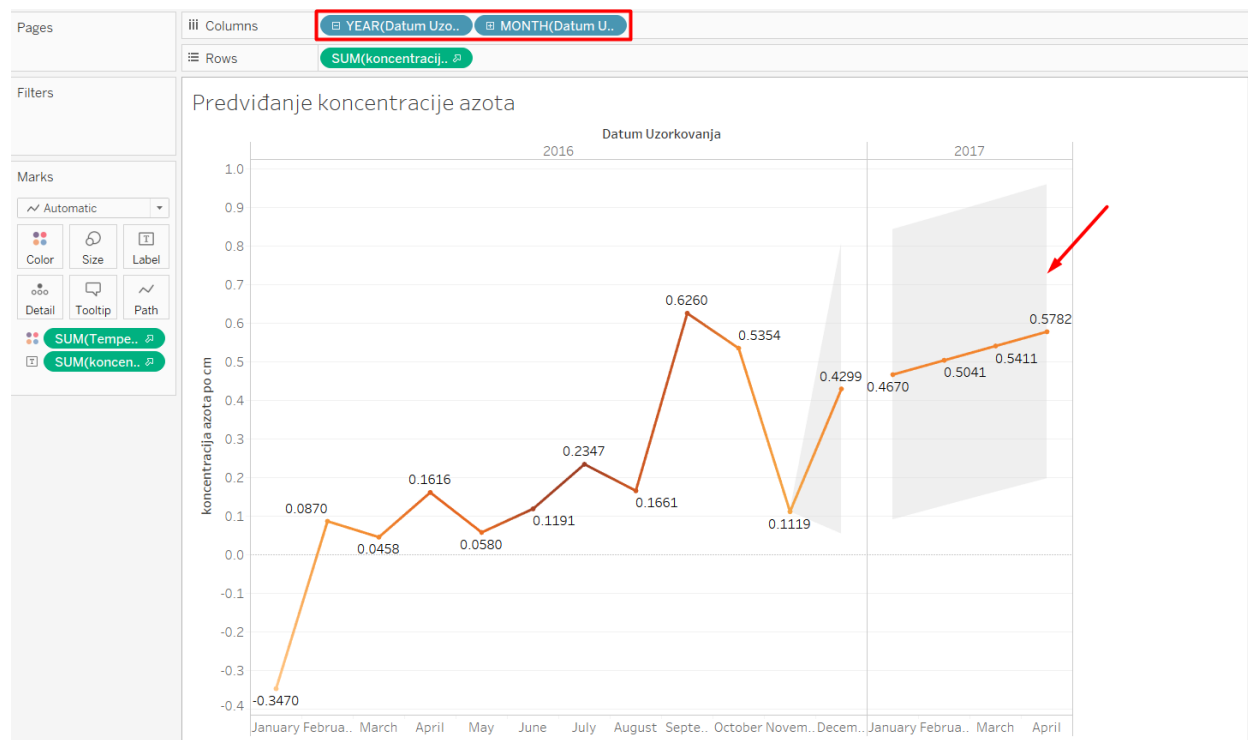
Главни *toolbar* садржи “дугмиће” за сортирање величина (од највеће вредности ка најмањој и обрнуто), за ротацију редова колона итд.

Додатне функционалности биће ближе објашњене кроз практичне примере

## Анализа података и додатни појмови

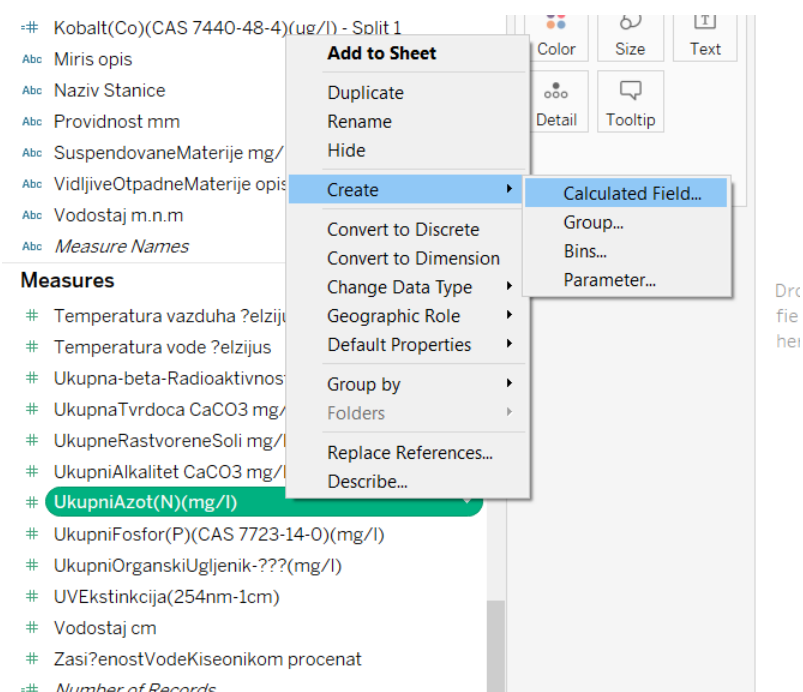
Пре свега битно је да се напомене да се у овом делу неће улазити у проблематику са стручне стране везано за квалитет воде и њен састав због недовољне компетентности. Циљ је да се најбоље могуће покажу функционалности које има овај програм. Текст са сликама и праћен објашњењима биће подељен по логичким јединицама.

## Drill down техника предвиђање и рачунање изведених мера



За потребе прегледа по месецима и приказивања предвиђања кретања одређене величине у првом кварталу наредне године коришћена је **“drill down”** техника. Она не представља ништа друго до хијерархијске декомпозиције исте димензије. Дакле, у овом случају временска димензија (датум узроковања) се може приказивати са што више детаља (од године до дана) кликом на ознаку + на пилулици (горњи део слике). Ви можете и сами креирати хијерархијски преглед простим превлачењем једне димензије до друге у левом делу *sheet*-а где се налазе све мере и димензије. Користан пример је хијерархија просторне димензије (континент-држава-град).

Поред већ постојећих мера могу се дефинисати и нове које представљају одређену рачуницу која укључује више њих. То су **изведене мере** (*calculated fields*) и на наредној слици је приказано како се креира нова формула. У датом примеру изведена мера се назива концентрација азота по цм и добијена је рачуницом:  $[\text{UkupniAzot(N)(mg/l)}]/[\text{Vodostaj cm}]$ .



Поред таба *Data* постоји још један *Analytics* који садржи опцију **Forecast** (предвиђање). *Drag and drop* техником са већ изабраним димензијама и мерама добија се предвиђање кретања вредности концентрације азота по цм.

## Сетови и агрегиране мере

Како се за мере долази до опције *calculated fields* на исти начин се за димензије долази до креирања **сетова**. Најчешће се креирају по неком критеријуму који зависи од потреба анализе. У овом случају креиран је сет за 5 највише рангираних станица по садржају CO2 елемента.

Name: top 5 stanica po CO2

General Condition Top

None

By field:

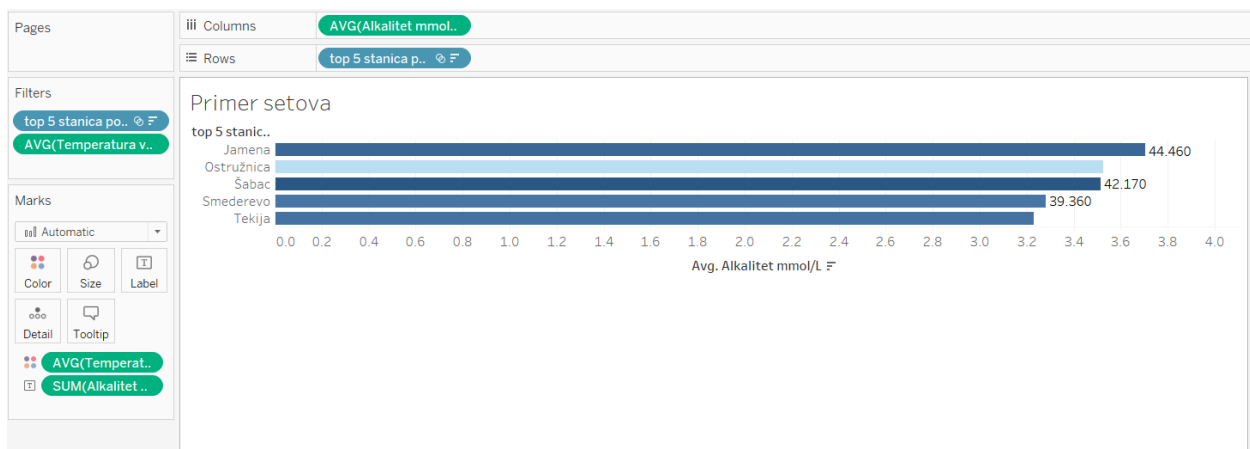
Top 5 by

SlobodniCO2 mg/l Average

By formula:

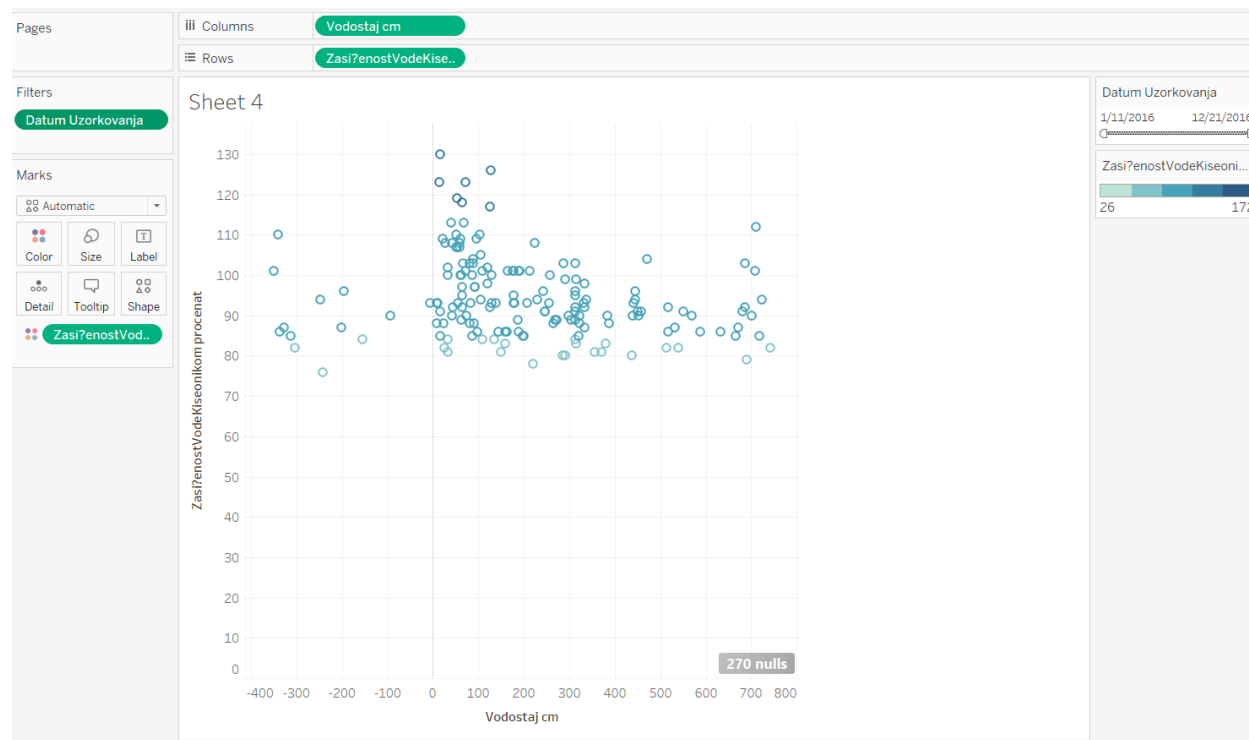
Top 10 by

Битно је напоменути да када се дефинисан сет изабере као димензија за предстојећу визуелизацију је неопходно левим кликом на “пилулицу” означити опцију *Show members in set*. Визуелизација за приказ овог сета у односу на меру *alkalitet mmol/L* и уз коришћење додатних филтера изгледа овако.



Поред сетова могу се креирати и групе одређених димензија, а ознака је спајалица. Групе су корисне нпр. када поседујемо информације у виду назива београдских општина (Нови Београ, Палилула, Чукарица) и све њих уврстимо у групу Београд.

**Агрегиране мере** представљају меру у виду суме, просека и сл. за сваки појединачни ред. До проблема долази када желимо да прикажемо одређену меру користећи визуелизацију *scatter plot* тј. тачкасти дијаграм. У том случају ће мера бити агрегирана у једну тачку И такав дијаграм неће бити ни од какве информативне значајности за нашу анализу. Опцијом на менију *Analysis/Aggregated measures* добијате следећу визуелизацију.

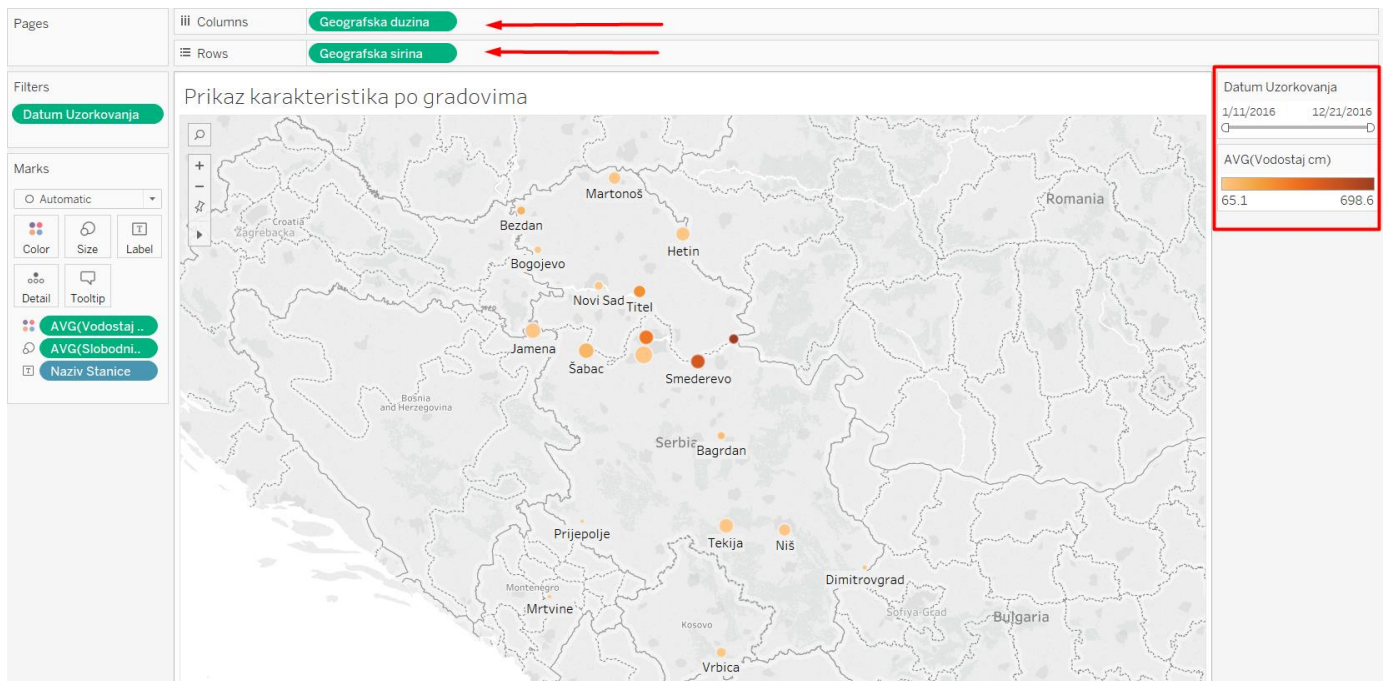


## Комбиновање извора података и географске димензије

Кликом на опцију у доњем левом углу *Data Source* у сваком тренутку се можемо вратити на почетни прозор и укључити још неки извор података. С обзиром да у оригиналном скупу података не постоји ниједна географска одредница позната овом софтверу направљен је Excel документ који садржи називе станица и њихове координате. Примарни *data set* и новонастали спојени су преко атрибута Назив места, а затим су у табеларном приказу варијабле географска ширина и географска дужина са обичне димензије промењене на географску димензију (*latitude i longitude*).

На наредној слици се види практична примена интерпретације ових димензија.





При коришћењу овог графикана географска дужина обавезно мора бити у пољу колоне, а ширина у пољу редова ради тачног локацијског приказа.

Као што се може видете тамнија боја представља висок водостај, а величина симбола (круг) концентрацију Slobodnog CO2. Све ове мере по различитим територијама могу се посматрати кроз време померањем курсора на филтер-у са десне стране.

## Dashbord i Story

Када сте подесили величину дасхборда потребно је изабрати опцију *Floating* како би се положај *sheet-ова* одређивао по личном нахођењу. Након тога дуплим кликом можете изабрати визуелизације, а пракса је да то буду оне које имају одређене заједничке мере или димензије због њеног упоредног приказа.

На главном менију избором опције *Dashbord/Action* можете дефинисати тзв. акцију за актуелни *dashboard*. У датом примеру коришћена је *Highlight/on hover* акција за паралелни приказ неког града на свим графиконима само преласком миша преко њега.

Са леве стране сваког *sheet-а* налазе се опције за њихово манипулисање респективно наведене – уклони *sheet*, иди на *sheet*, прикажи филтере.

Dashboard Layout

Device Preview

Size

Generic Desktop (1366 x 768)

Sheets

- Prosečna temperatu...
- Vodostaj po meseci...
- Ukupni alkalinitet po ...
- Zasićenost kiseonik...
- Koncentracija azota ...
- Predviđanje koncent...
- Primer setova
- Prikaz karakteristika...

Objects

- Horizontal
- Image
- Vertical
- Web Page
- Text
- Blank

Tiled Floating

Show dashboard title

### Geografski pregled

#### Ukupni alkalinitet po stanicama

Naziv Stanice	Ukupni Alkalinitet CaCO3 mg/l
Vrbica	4,665
Hetin	2,602
Mrtvine	2,208
Jamena	2,118
Dimitrovgrad	2,118
Niš	2,118
Ostružnica	2,118
Sabac	2,109
Bagrdan	1,918
Bogojevo	1,918
Bezdan	1,918
Smederevo	1,918
Novi Sad	1,918
Tekija	1,918
Zemun	1,918
Banatska Palanka	1,668
Prijepolje	1,668
Titel	1,404
Martonos	1,404

Month of Datum Uzorkovanja

- (All)
- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December

Prikaz karakteristika po gradovima

Avg. Temperatura vode izljud: 14.450

Ukupni Alkalinitet CaCO3 mg/l: 2,109

Avg. Temperatura va...: 9.375 - 15.575

Datum Uzorkovanja: 1/11/2016 - 12/21/2016

top 5 stanic...

Station	Avg. Alkalinitet mmol/L
Jamena	44.460
Ostružnica	42.170
Sabac	39.760
Smederevo	38.760
Tekija	38.760

Напокон смо дошли до финалне тачке сваке анализе, а то је креирање *Story-ja* који представља сумарни приказ најбитнијих запажања и информација. Његови чиниоци могу бити појединачни графיקони или “табле”, које су и изабране у овом случају.

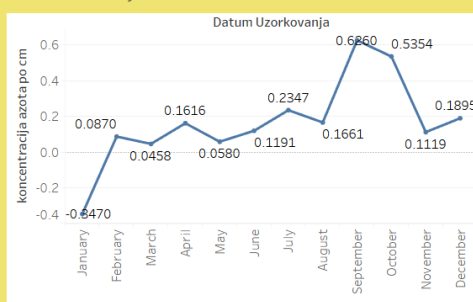
Као и код *dashboard*-а дуплим кликом се бирају визуелизације изнад којих ће стајати кратак закључак И може се подесити величина поља. Престојећа слика показује једну страницу нашег *Story-ja*.

## Kvalitet vode u 2016. godini

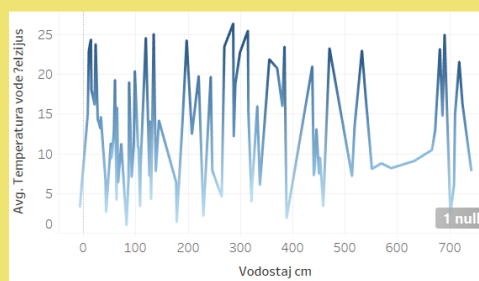
Koncentracija azota je najviša tokom avgusta što je u direktonoj vezi sa temperaturom koja utiče na vodostaj. Što je vodostaj niži, koncentracija je veća. Stoga u avgustu iznosi 0.6260.

Uporedni prikaz mesta po različitim karakteristikama nam pruža uvid u kvalitet vode kroz različite vremenske periode.

### Koncentracija azota kroz mesece



### Prosečna temperatura vode na različitim dubinama



### Predviđanje koncentracije azota

