

Process Book

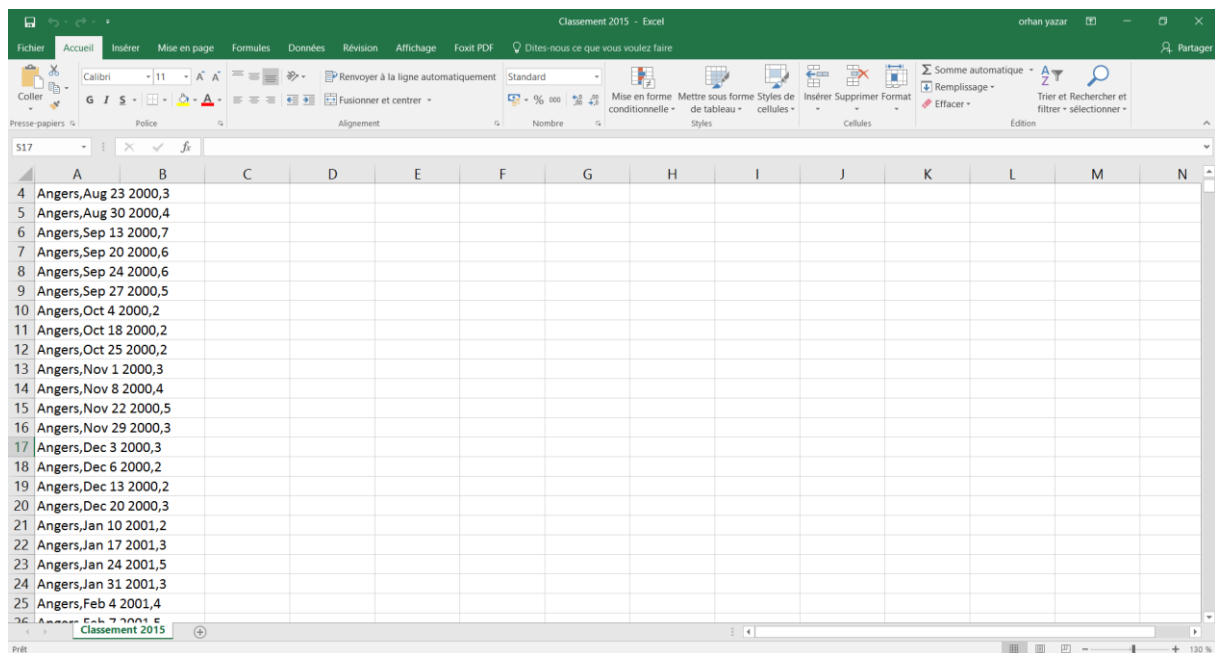
Après le choix du sujet, qui est la visualisation de classements, en l'occurrence ici des classements de football. Nous avons décidés de visualiser au départ les classements de ligue 1 sur les 30 dernières années. Mais finalement nous nous sommes restreint à une saison seulement du fait de la difficulté et du temps que nous prendrait d'une part l'acquisition d'autant de donnée et d'autre part le traitement de ces données.

Nous avons donc commencé par écrire un code en python pour du web scrapping pour scrapper une vingtaine de saison de ligue 1.

```
v=np.r_[5:60:1]
for int in v:
    vv=str(int)
    vec=[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38]
    dico = {}
    for i in vec:
        a = str(i)
        url1 = 'http://www.lfp.fr/ligue1/competitionPluginClassement/loadClassement?sai='+vv+'&journee1=1&journee2='+a+'&cat=Gen.'
        buffer = StringIO()
        c = pycurl.Curl()
        c.setopt(c.URL,url1)
        c.setopt(pycurl.HTTPHEADER, ['X-Requested-With: XMLHttpRequest', 'Content-Type: application/x-www-form-urlencoded'])
        c.setopt(c.WRITEDATA, buffer)
        c.perform()
        c.close()
        bodi = buffer.getvalue()
        time.sleep(1)
        soup = BeautifulSoup(bodi, 'html.parser')
        lignes = soup.findAll('tr')
        for l in lignes:
            if (len(l.findAll('td'))==11):
                team = l.findAll('td')[2].text.replace(" ", "").replace("\n", "").replace("\t", "").encode('ascii', 'ignore') + \
                    l.findAll('td')[0].text
                teamf = re.sub(r'([1-9].*|\s)$', r'\1', team)
                teamff = teamf.split(',')
                if teamff[0] in dico.keys():
                    dico[teamff[0]].append([teamff[1]])
                else:
                    dico[teamff[0]] = [[teamff[1]]]
```

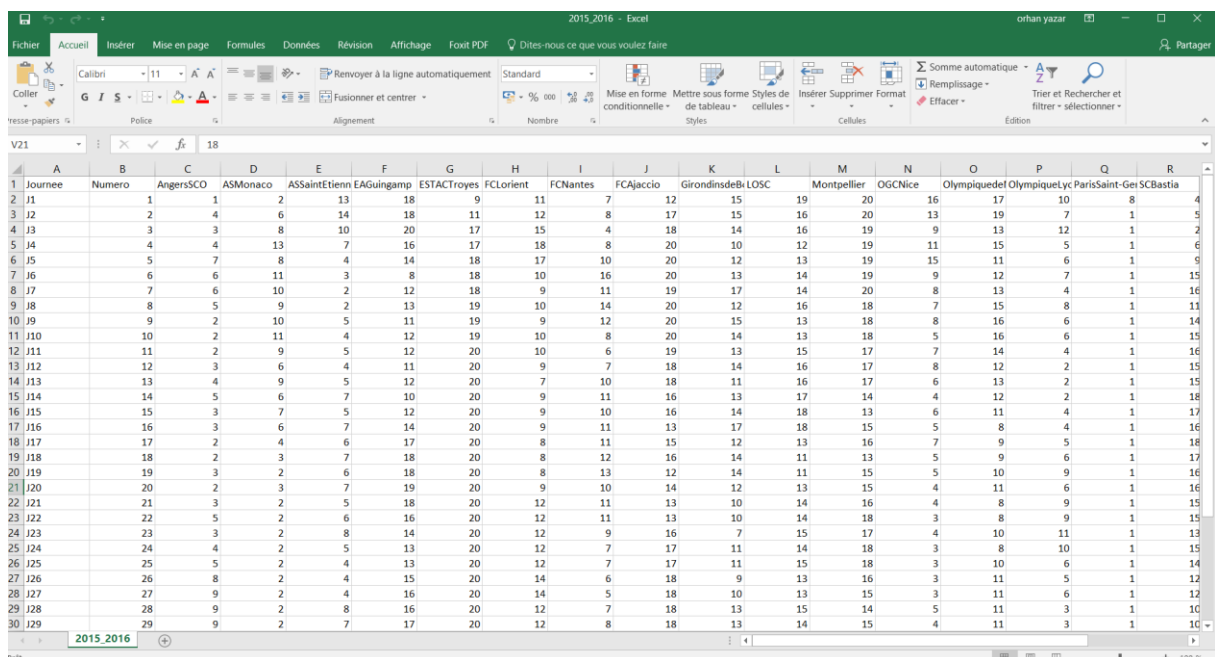
Le problème était le traitement des données une fois acquises.

Voici ce qu'on a obtenu après le scrapping.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
4	Angers,Aug 23 2000,3													
5	Angers,Aug 30 2000,4													
6	Angers,Sep 13 2000,7													
7	Angers,Sep 20 2000,6													
8	Angers,Sep 24 2000,6													
9	Angers,Sep 27 2000,5													
10	Angers,Oct 4 2000,2													
11	Angers,Oct 18 2000,2													
12	Angers,Oct 25 2000,2													
13	Angers,Nov 1 2000,3													
14	Angers,Nov 8 2000,4													
15	Angers,Nov 22 2000,5													
16	Angers,Nov 29 2000,3													
17	Angers,Dec 3 2000,3													
18	Angers,Dec 6 2000,2													
19	Angers,Dec 13 2000,2													
20	Angers,Dec 20 2000,3													
21	Angers,Jan 10 2001,2													
22	Angers,Jan 17 2001,3													
23	Angers,Jan 24 2001,5													
24	Angers,Jan 31 2001,3													
25	Angers,Feb 4 2001,4													

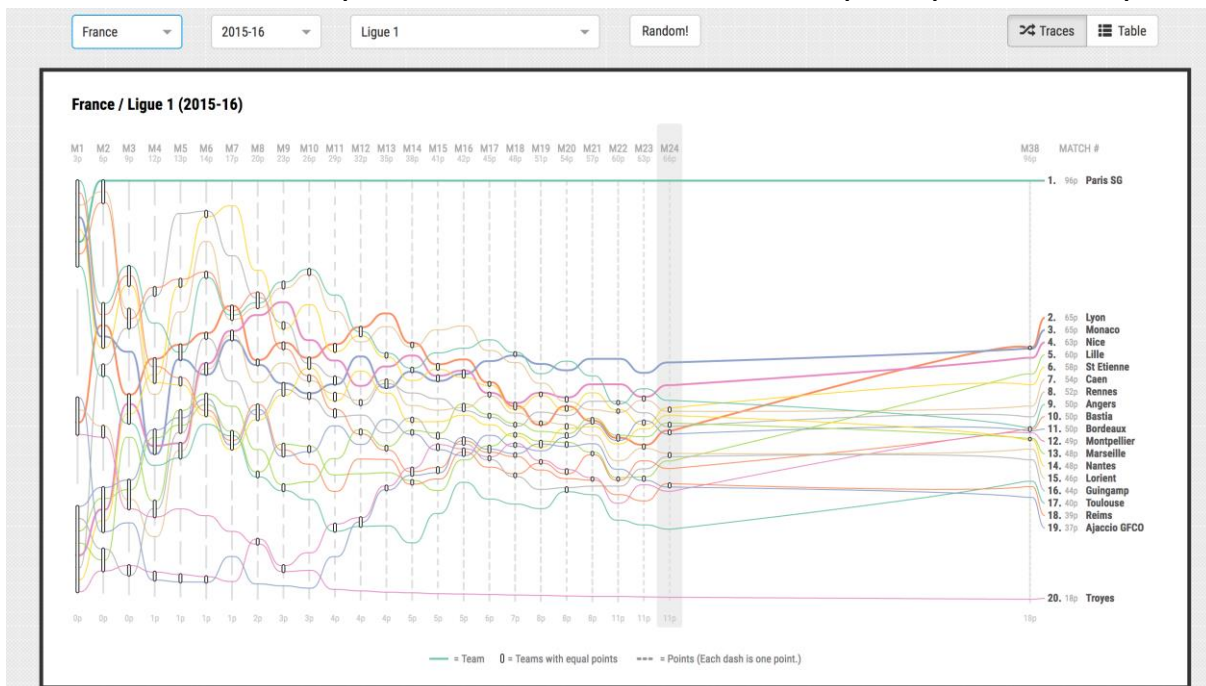
Au premier abord, ce format ne semble pas très utile et nous l'avons modifié en le format suivant.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Journee	Numero	AngersSCO	ASMonaco	ASSaintEtienne	EAGuingamp	ESTACTroyes	FLorient	FCNantes	FCAjaccio	GirondinsdeB	LOSC	Montpellier	OGCNice	OlympiqueLyc	OlympiqueLyc	ParisSaint-Ge	SCBastia
2	J1	1	2	13	18	9	11	7	12	15	19	20	16	17	10	8	4	
3	J2	2	4	6	14	18	11	12	8	17	15	16	20	13	19	7	1	5
4	J3	3	8	10	20	17	15	4	18	14	16	19	9	13	12	1	2	
5	J4	4	4	13	7	16	17	18	8	20	10	12	19	11	15	5	1	6
6	J5	5	7	8	4	14	18	17	10	20	12	13	19	15	11	6	1	9
7	J6	6	11	3	8	18	10	16	20	13	14	19	9	12	7	1	15	
8	J7	7	6	10	2	12	18	9	11	19	17	14	20	8	13	4	1	16
9	J8	8	5	9	2	13	19	10	14	20	12	16	18	7	15	8	1	11
10	J9	9	2	10	5	11	19	9	12	20	15	13	18	8	16	6	1	14
11	J10	10	2	11	4	12	19	10	8	20	14	13	18	5	16	6	1	15
12	J11	11	2	9	5	12	20	10	6	19	13	15	17	7	14	4	1	16
13	J12	12	3	6	4	11	20	9	7	18	14	16	17	8	12	2	1	15
14	J13	13	4	9	5	12	20	7	10	18	11	16	17	6	13	2	1	15
15	J14	14	5	6	7	10	20	9	11	16	13	17	14	4	12	2	1	18
16	J15	15	3	7	5	12	20	9	10	16	14	18	13	6	11	4	1	17
17	J16	16	3	6	7	14	20	9	11	13	17	18	15	5	8	4	1	16
18	J17	17	2	4	6	17	20	8	11	15	12	13	16	7	9	5	1	18
19	J18	18	2	3	7	18	20	8	12	16	14	11	13	5	9	6	1	17
20	J19	19	3	2	6	18	20	8	13	12	14	11	15	5	10	9	1	16
21	J20	20	2	3	7	19	20	9	10	14	12	13	15	4	11	6	1	16
22	J21	21	3	2	5	18	20	12	11	13	10	14	16	4	8	9	1	15
23	J22	22	5	2	6	16	20	12	11	13	10	14	18	3	8	9	1	15
24	J23	23	3	2	8	14	20	12	9	16	7	15	17	4	10	11	1	13
25	J24	24	4	2	5	13	20	12	7	17	11	14	18	3	8	10	1	15
26	J25	25	5	2	4	13	20	12	7	17	11	15	18	3	10	6	1	14
27	J26	26	8	2	4	15	20	14	6	18	9	13	16	3	11	5	1	12
28	J27	27	9	2	4	16	20	14	5	18	10	13	15	3	11	6	1	12
29	J28	28	9	2	8	16	20	12	7	18	13	15	14	5	11	3	1	10
30	J29	29	9	2	7	17	20	12	8	18	13	14	15	4	11	3	1	10

Finalement, lorsque nous avons commencé à coder en javascript, nous avons préférés utiliser le premier format que nous avons trouvé plus simple d'utilisation.

Concernant la visualisation, nous nous sommes directement orienté vers une visualisation linechart et monsieur Vuillemot nous a donné un lien ou on pouvait trouver un exemple pour s'inspirer.



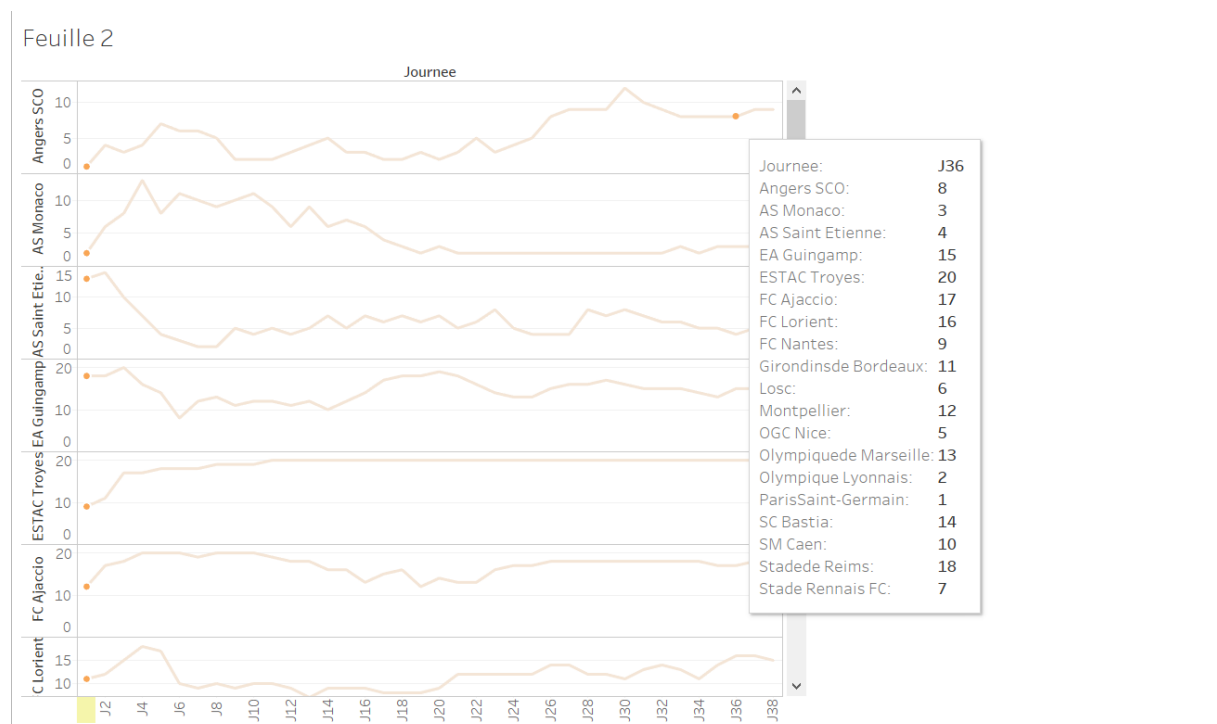
Evidemment, nous savions que notre visualisation ne serait pas aussi évoluée. Bien sûr, il nous fallait une valeur ajoutée à la visualisation car un linechart sur l'évolution du classement ne suffirait pas, et nous avons immédiatement eu l'idée d'ajouter des prédictions. Pour cela, nous avons beaucoup hésité concernant le modèle à utiliser et nous ne pouvions pas nous permettre de les tester toutes car il y en a beaucoup trop nous avons donc décidé d'utiliser une méthode AFD pour la simple et bonne raison que cette une nouvelle méthode statistique que nous venions d'apprendre en analyse factorielle. Nous avons utilisés le logiciel R pour cela, voici le code que nous avons écrit.

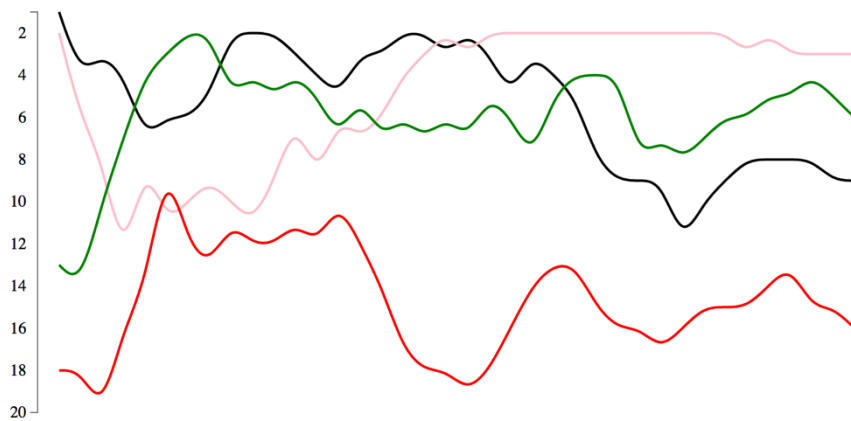
```

1 setwd('C:/Users/letur/Desktop/DataViz/Projet_DataViz')
2 table=read.csv('C:/Users/letur/Desktop/DataViz/Projet_DataViz/log.csv',header=TRUE,sep=";")
3 table[1][table[1]==1]
4 attach(table)
5 summary(table)
6 reg=glm(Resultat~Classement,family=binomial(link='logit'))
7 summary(reg)
8 confusion.glm <- function(data, model) {
9   prediction <- ifelse(predict(model, data, type='response') > 0.5, TRUE, FALSE)
10  confusion <- table(prediction, as.logical(model$y))
11  confusion <- cbind(confusion, c(1 - confusion[1,1]/(confusion[1,1]+confusion[2,1]), 1 - confusion[2,2]/(confusion[2,2]+confusion[1,2])))
12  confusion <- as.data.frame(confusion)
13  names(confusion) <- c('FALSE', 'TRUE', 'class.error')
14  confusion
15 }
16 confusion.glm(table,reg)
17

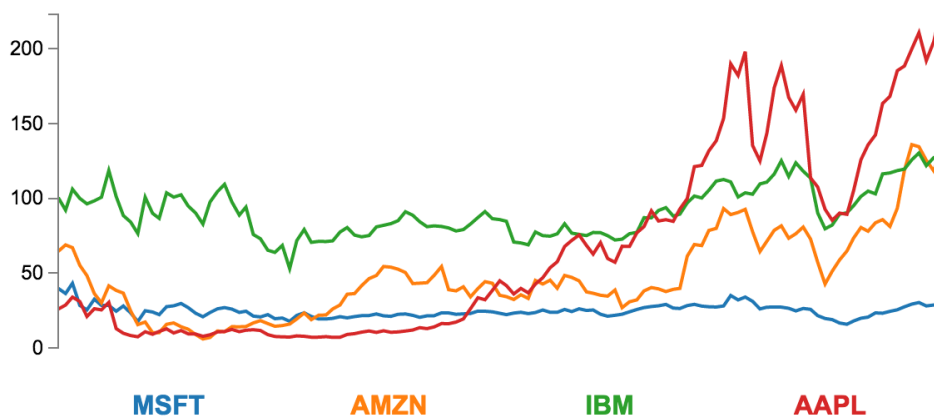
```

On a commencé notre visualisation avec la réalisation de line charts simples. Très rapidement après l'implémentation de plusieurs courbes on se rend compte que la visualisation sature très vite (plus de 5 courbes).



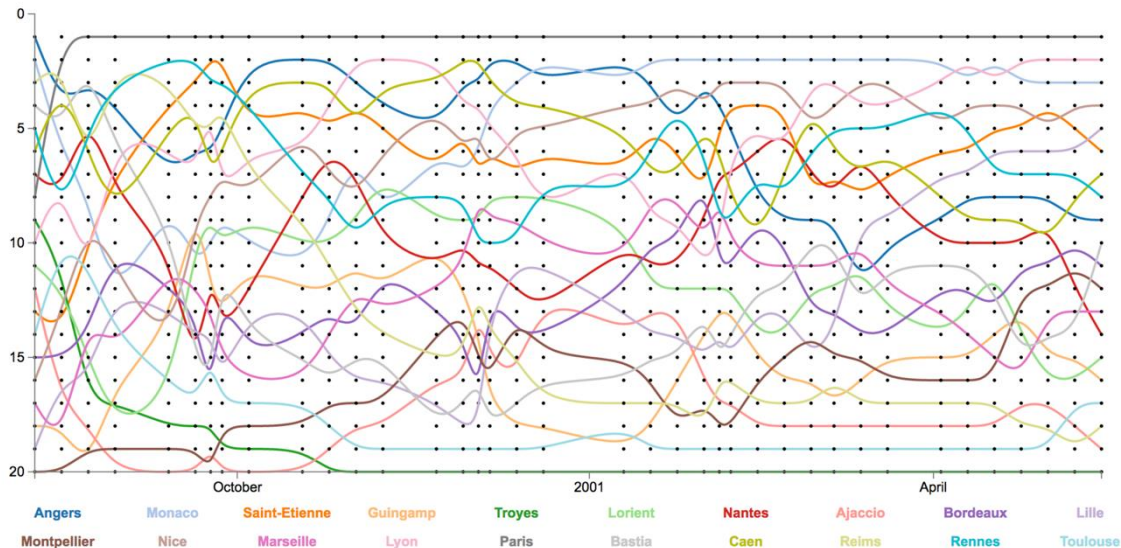


On a donc décidé d'insérer une légende interactive inspiré du graphe suivant. L'implémentation de cette légende s'avérera être essentiel pour permettre une bonne lisibilité et une comparaison facile.



Pour l'étape suivante on a essayé de travailler avec des fonctions d'interpolation, notamment la fonction `.interpolate(«basis»)`. Cette fonction nous a posée des problèmes concernant les points associés. En effet quand les courbes sont interpolées elles ne passent plus précisément par les points. Il faut donc coordonner les courbes et les

points ce que nous n'avons pas réussi à réaliser. Nous sommes donc revenus à un modèle plus classique comme style de courbe.



On a pu obtenir des informations concernant ces interpolations sur le site stackoverflow.

share edit delete flag

edited Dec 26 '16 at 23:29



altocumulus

6,615 ● 8 ● 17 ● 40

asked Dec 26 '16 at 11:02



Florian Robinet

1

This code is working, and the dots are not repeating. Check this fiddle, using your code with bogus data: jsfiddle.net/gerardofurtado/a5eegveL. I'll try to guess what's happening here: the teams (France league?) in your CSV are ranked from 1 to 20, for each round. So, for each date, you have (not the same) dots going from 1 to 20. That's why the dots seem to be repeating, but they are not. This will be even more clear if you remove the `interpolate("basis")`. – Gerardo Furtado Dec 26 '16 at 14:07

You are right, thank you very much, when i remove `interpolate("basis")` it works well, so how could i link `interpolate("basis")` and `scatterplot` ? – Florian Robinet Dec 26 '16 at 15:58

Try different interpolators with different tensions: github.com/d3/d3-3.x-api-reference/blob/master/... – Gerardo Furtado Dec 27 '16 at 0:29

[add a comment](#)