Hi, I’m Ben and these are the groups I care about

@plzbeemyfriend

@RLadiesGlobal

@R_LGBTQ

@rOpenSci

Data Manipulation With

Vectors

```
> letters
 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m"
[14] "n" "o" "p" "q" "r" "s" "t" "u" "v" "w" "x" "y" "z"
> 1:length(letters)
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
[19] 19 20 21 22 23 24 25 26
> !!1:length(letters) %% 2
 [1]  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE
[10] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE
[19]  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE
```

Data Frames (tables with rows and columns)

	characters	numerics	logicals
1	a	1	TRUE
2	b	2	FALSE
3	c	3	TRUE
4	d	4	FALSE
5	e	5	TRUE
6	f	6	FALSE
7	g	7	TRUE
8	h	8	FALSE
9	i	9	TRUE
10	j	10	FALSE
11	k	11	TRUE
12	l	12	FALSE
13	m	13	TRUE
14	n	14	FALSE
15	o	15	TRUE
16	p	16	FALSE

Tidy Data

Journal of Statistical Software
MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

Tidy Data

Hadley Wickham
RStudio

Abstract

A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualise, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table. This framework makes it easy to tidy messy datasets because only a small set of tools are needed to deal with a wide range of un-tidy datasets. This structure also makes it easier to develop tidy tools for data analysis, tools that both input and output tidy datasets. The advantages of a consistent data structure and matching tools are demonstrated with a case study free from mundane data manipulation chores.

Keywords: data cleaning, data tidying, relational databases, R.

1. Each variable forms a column
2. Each observation forms a row
3. Each type of observational unit forms a table

Tidy Data - Shape of Data Matters

religion	<$10k	$10-20k	$20-30k	$30-40k	$40-50k	$50-75k
Agnostic	27	34	60	81	76	137
Atheist	12	27	37	52	35	70
Buddhist	27	21	30	34	33	58
Catholic	418	617	732	670	638	1116
Don't know/refused	15	14	15	11	10	35
Evangelical Prot	575	869	1064	982	881	1486
Hindu	1	9	7	9	11	34
Historically Black Prot	228	244	236	238	197	223
Jehovah's Witness	20	27	24	24	21	30
Jewish	19	19	25	25	30	95

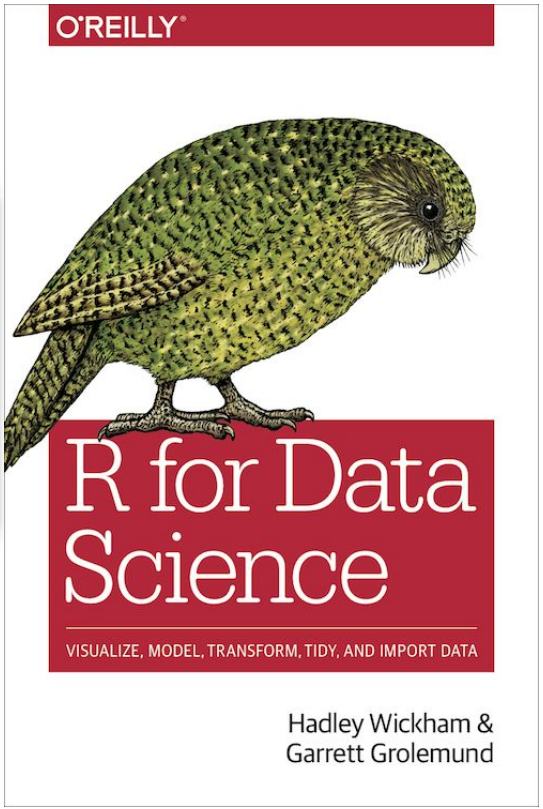
religion	income	freq
Agnostic	<$10k	27
Agnostic	$10-20k	34
Agnostic	$20-30k	60
Agnostic	$30-40k	81
Agnostic	$40-50k	76
Agnostic	$50-75k	137
Agnostic	$75-100k	122
Agnostic	$100-150k	109
Agnostic	>150k	84
Agnostic	Don't know/refused	96

Tidy Data - Content of Cells Matter

country	year	rate
Afghanistan	1999	**745** / 19987071
Afghanistan	2000	**2666** / 20595360
Brazil	1999	**37737** / 172006362
Brazil	2000	**80488** / 174504898
China	1999	**212258** / 1272915272
China	2000	**213766** / 1280428583

country	year	cases	population
Afghanistan	1999	**745**	19987071
Afghanistan	2000	**2666**	20595360
Brazil	1999	**37737**	172006362
Brazil	2000	**80488**	174504898
China	1999	**212258**	1272915272
China	2000	**213766**	1280428583

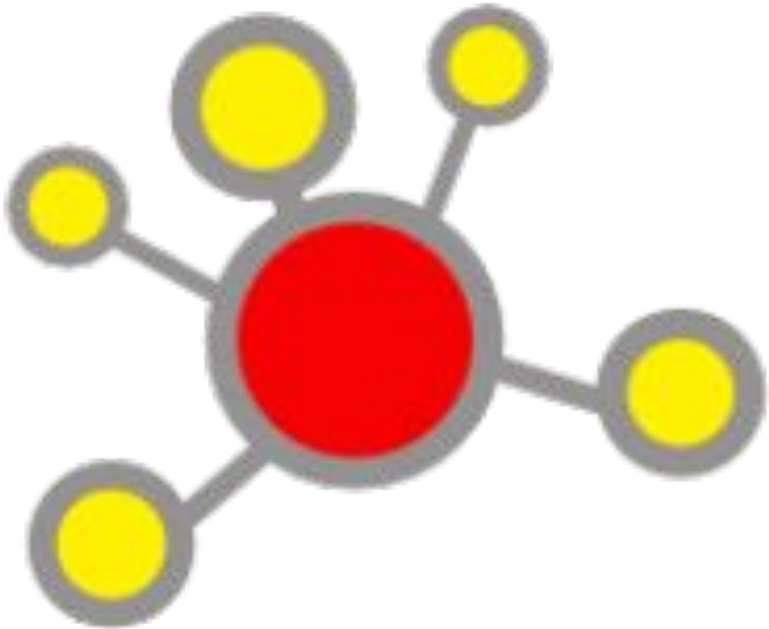
Tidy Data - Tools & Resources



http://r4ds.had.co.nz

@hadleywickham

Goal - Format Data to Create Graphs with igraph



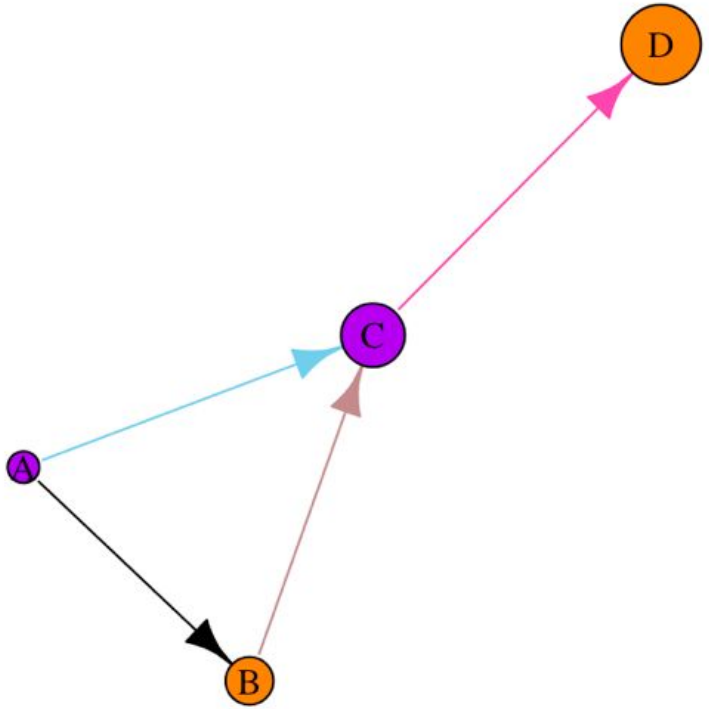
igraph::graph_from_data_frame()

Edge List

Source	Target	Color
A	B	black
A	C	skyblue
B	C	rosybrown
C	D	hotpink

Node List

Name	color	Size
A	purple	10
B	darkorange	15
C	purple	20
D	darkorange	25



Resources For igraph

http://kateto.net/tutorials/

@Ognyanova

Network Science Tutorials

This page contains information about the most recent versions of several network tutorials that I have developed and frequently update. The tutorials come from workshops and invited talks I give for students, colleagues, and computationally curious bystanders. You can also find some of these materials (and other interesting bits and pieces) on my GitHub page.

If you find the materials on this site to be useful, please cite them in your work. This helps me (and the computational research community) to make the case that the open publishing of digital materials, data, and code is a meaningful academic contribution.

If you want to invite me to give a talk or a workshop at your institution, email workshop@ognyanova.net.

Search ...

Katherine Ognyanova

Asst. Prof. Rutgers SC&I
Postdoc at the Lazer Lab: Northeastern & Harvard.
PhD in Communication: USC Annenberg School.

Where Can We Get Data?

Where Can We Get Data?

aRxiv::arxiv_search()

QUERY arXiv for: **'abs:social AND abs:network'**

	id	title	summary	published	primary_category	author	affiliation
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	http://…	"Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	Mark …	Univ. of Wis…
2	http://…	"Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	E. An…	Iowa State U…
3	http://…	"Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	Dan A…	Iowa State U…
4	http://…	Computer …	" A simula…	1998-05-…	cond-mat.stat-mec…	D. He…	""
5	http://…	Computer …	" A simula…	1998-05-…	cond-mat.stat-mec…	P. Mo…	""
6	http://…	Computer …	" A simula…	1998-05-…	cond-mat.stat-mec…	F. Sc…	""
7	http://…	Dynamics …	" A new ap…	1998-07-…	adap-org | http:/…	Makot…	""
8	http://…	Dynamics …	" A new ap…	1998-07-…	adap-org | http:/…	Takas…	""
9	http://…	"A Delay …	" In heter…	1998-09-…	cs.NI | http://ar…	R. Ja…	""
10	http://…	"Autocata…	" This pap…	1999-01-…	adap-org | http:/…	L. Ga…	""

... with 19,417 more rows

Collaboration Networks

	id	title	summary	published	primary_category	author	affiliation
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	http://…	Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	Mark …	niv. of Wis…
2	http://…	Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	E. An…	owa State U…
3	http://…	Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	Dan A…	owa State U…
4	http://…	omputer …	" A simula…	1998-05-…	cond-mat.stat-mec…	D. He…	"
5	http://…	omputer …	" A simula…	1998-05-…	cond-mat.stat-mec…	P. Mo…	"
6	http://…	omputer …	" A simula…	1998-05-…	cond-mat.stat-mec…	F. Sc…	"
7	http://…	ynamics …	" A new ap…	1998-07-…	adap-org | http:/…	Makot…	"
8	http://…	ynamics …	" A new ap…	1998-07-…	adap-org | http:/…	Takas…	"
9	http://…	A Delay …	" In heter…	1998-09-…	cs.NI | http://ar…	R. Ja…	"
10	http://…	Autocata…	" This pap…	1999-01-…	adap-org | http:/…	L. Ga…	"

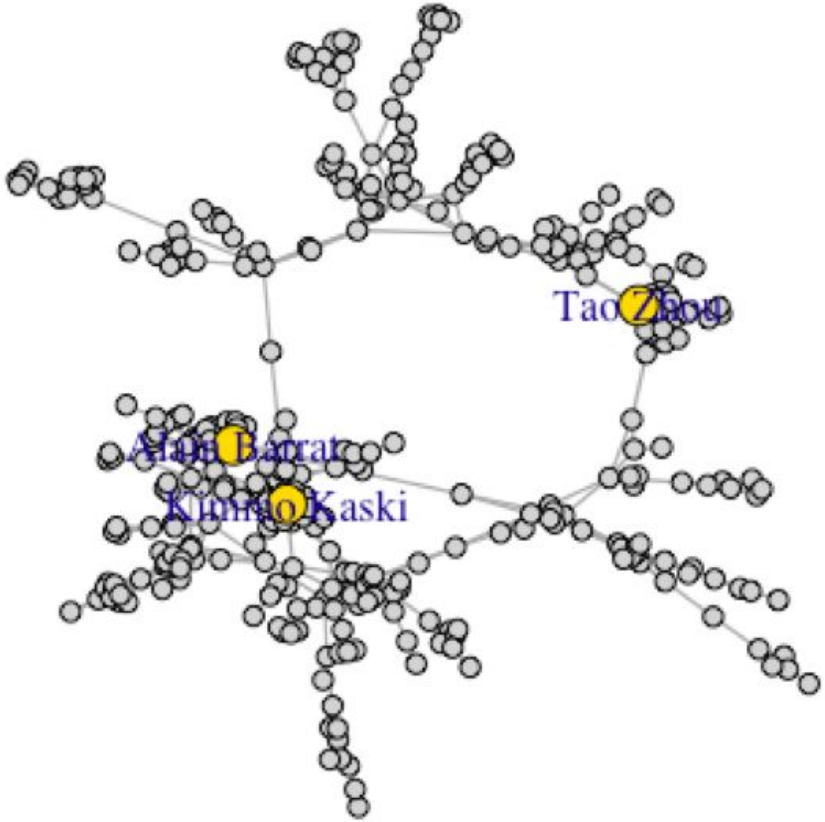
… with 15,417 more rows

Working With Publication Data

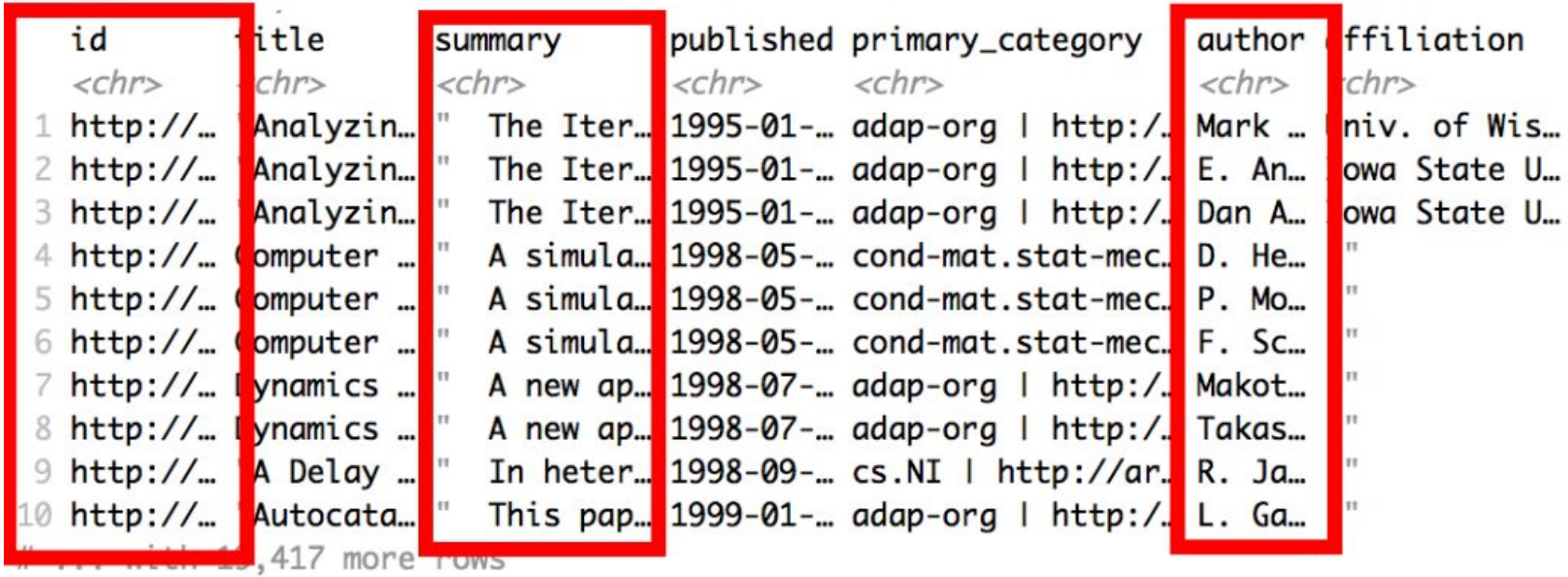
	from	to	n
	<chr>	<chr>	<int>
1	Piotr Bródka	Przemysław Kazienko	19
2	Attila Szolnoki	Matjaz Perc	18
3	Anastasios Noulas	Cecilia Mascolo	14
4	Alain Barrat	Ciro Cattuto	13
5	Alessandro Flammini	Filippo Menczer	13
6	Kristina Lerman	Rumi Ghosh	13
7	Emilio Ferrara	Giacomo Fiumara	12
8	Emilio Ferrara	Pasquale De Meo	12
9	János Kertész	Kimmo Kaski	11
10	Ali Pinar	C. Seshadhri	10

... with 25,186 more rows

Largest Component of Collaboration Network



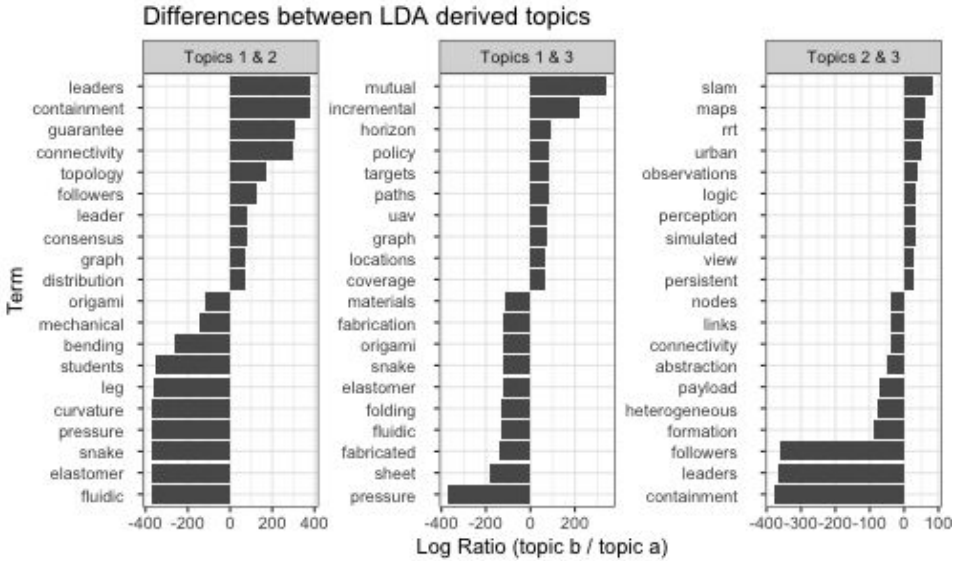
Modelling Text Data (LDA)



	id	title	summary	published	primary_category	author	affiliation
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	http://…	Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	Mark …	niv. of Wis…
2	http://…	Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	E. An…	owa State U…
3	http://…	Analyzin…	" The Iter…	1995-01-…	adap-org | http:/…	Dan A…	owa State U…
4	http://…	omputer …	" A simula…	1998-05-…	cond-mat.stat-mec…	D. He…	"
5	http://…	omputer …	" A simula…	1998-05-…	cond-mat.stat-mec…	P. Mo…	"
6	http://…	omputer …	" A simula…	1998-05-…	cond-mat.stat-mec…	F. Sc…	"
7	http://…	ynamics …	" A new ap…	1998-07-…	adap-org | http:/…	Makot…	"
8	http://…	ynamics …	" A new ap…	1998-07-…	adap-org | http:/…	Takas…	"
9	http://…	A Delay …	" In heter…	1998-09-…	cs.NI | http://ar…	R. Ja…	"
10	http://…	Autocata…	" This pap…	1999-01-…	adap-org | http:/…	L. Ga…	"

… with 15,417 more rows

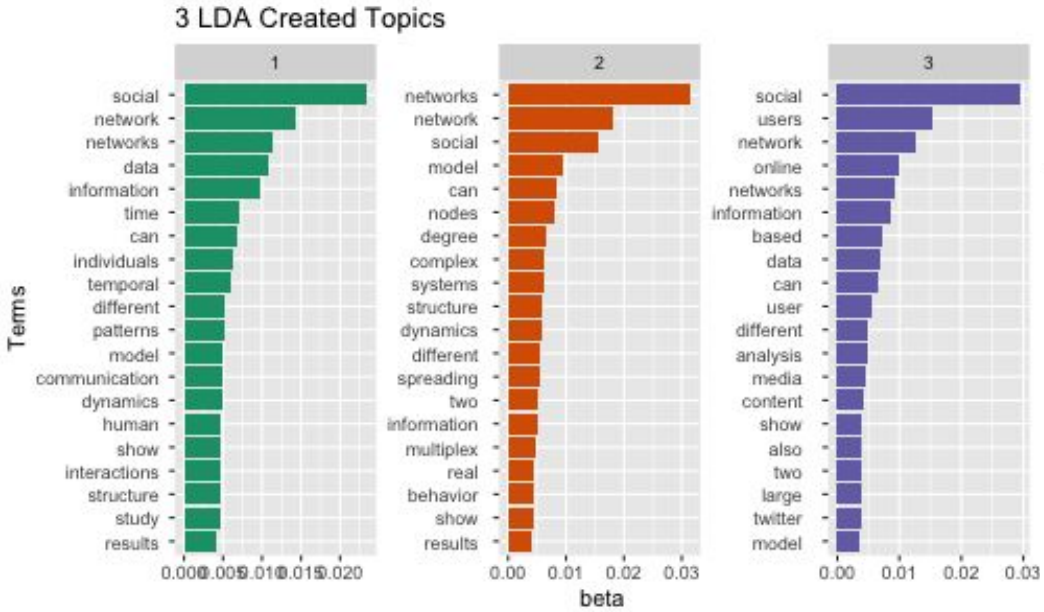
tidytext::unnest_tokens()



	id	word
	<chr>	<chr>
1	http://arxiv.org/abs/0909.4019v1	study
2	http://arxiv.org/abs/0909.4019v1	evolution
3	http://arxiv.org/abs/0909.4019v1	cooperation
4	http://arxiv.org/abs/0909.4019v1	prisoner's
5	http://arxiv.org/abs/0909.4019v1	dilemma
6	http://arxiv.org/abs/0909.4019v1	game
7	http://arxiv.org/abs/0909.4019v1	whereby
8	http://arxiv.org/abs/0909.4019v1	coevolutionary
9	http://arxiv.org/abs/0909.4019v1	rule
10	http://arxiv.org/abs/0909.4019v1	introduced

... with 56,527 more rows

topicmodels::LDA()



	topic	term	beta
	<int>	<chr>	<dbl>
1	1	ability	4.84e- 4
2	2	ability	1.25e- 4
3	3	ability	3.36e- 4
4	1	activity	1.21e- 3
5	2	activity	2.65e- 3
6	3	activity	2.39e- 3
7	1	adaptability	5.09e- 5
8	2	adaptability	7.48e-98
9	3	adaptability	1.05e- 4
10	1	algorithm	2.31e- 3

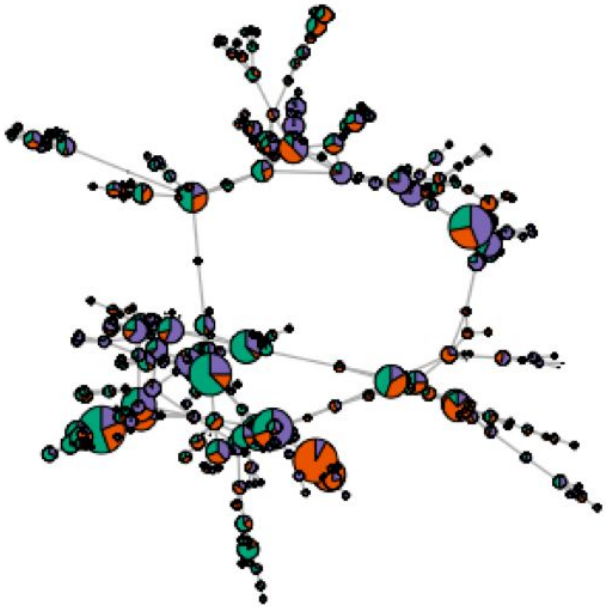
... with 18,959 more rows

	id	topic	gamma
	<chr>	<int>	<dbl>
1	http://arxiv.org/abs/0704.0744v1	1	0.999
2	http://arxiv.org/abs/0704.1589v1	2	0.692
3	http://arxiv.org/abs/0710.2595v1	2	0.734
4	http://arxiv.org/abs/0801.1896v1	2	0.998
5	http://arxiv.org/abs/0802.0047v2	2	0.998
6	http://arxiv.org/abs/0805.0512v2	1	0.999
7	http://arxiv.org/abs/0805.2976v3	1	0.714
8	http://arxiv.org/abs/0807.0750v4	2	0.618
9	http://arxiv.org/abs/0807.3089v4	1	0.552
10	http://arxiv.org/abs/0808.2195v1	3	0.727

... with 500 more rows

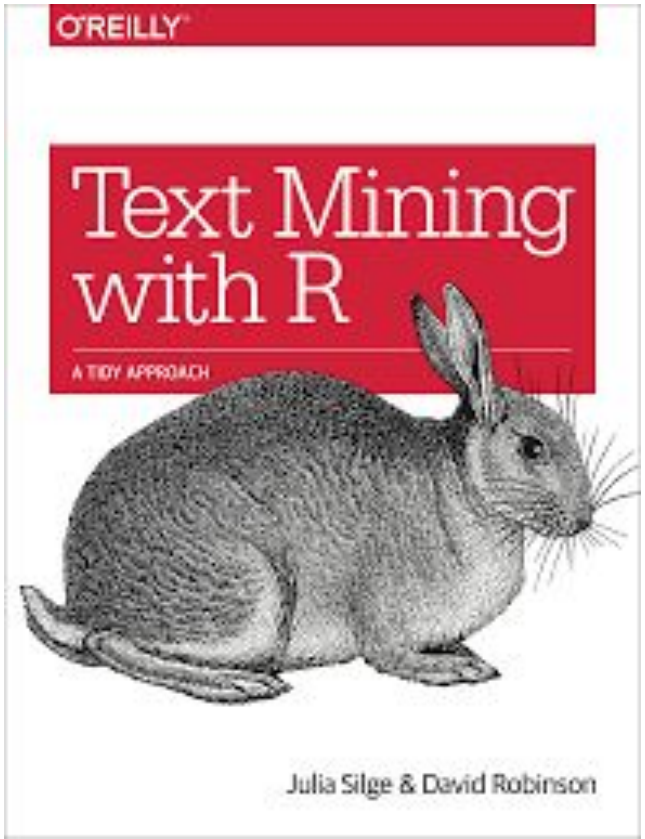
Incorporating LDA Topics to Networks

Largest Component Colored by Topics



name	pie	size	label	shape	pie.color
Attila Szolnoki	0, 17, 1	11.000000		pie	#1b9e77, #d95f02, #7570b3
Anastasios Noulas	2, 12, 1	9.333333		pie	#1b9e77, #d95f02, #7570b3
Alessandro Flammini	6, 6, 5	10.444444		pie	#1b9e77, #d95f02, #7570b3
Emilio Ferrara	15, 7, 5	16.000000		pie	#1b9e77, #d95f02, #7570b3
János Kertész	8, 0, 7	9.333333		pie	#1b9e77, #d95f02, #7570b3
Giacomo Fiumara	6, 3, 3	7.666667		pie	#1b9e77, #d95f02, #7570b3

Resources For Text



https://www.tidytextmining.com/

@juliasilge
@drob

Where Can We Get Data?

We Have Trip Information

	start	end	month	n
	<chr>	<chr>	<ord>	<int>
1	Jefferson Dr & 14th St SW	Jefferson Dr & 14th St SW	Apr	1212
2	Jefferson Dr & 14th St SW	Jefferson Dr & 14th St SW	Jun	1096
3	Lincoln Memorial	Jefferson Memorial	Jul	1047
4	Jefferson Dr & 14th St SW	Jefferson Dr & 14th St SW	Jul	1031
5	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Jul	982
6	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Apr	962
7	Lincoln Memorial	Jefferson Memorial	Aug	943
8	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Jun	916
9	Jefferson Dr & 14th St SW	Lincoln Memorial	Jun	850
10	Lincoln Memorial	Jefferson Memorial	Apr	837

... with 445,129 more rows

Similar to Edge List

	start	end	month	n
	<chr>	<chr>	<ord>	<int>
1	Jefferson Dr & 14th St SW	Jefferson Dr & 14th St SW	Apr	1212
2	Jefferson Dr & 14th St SW	Jefferson Dr & 14th St SW	Jun	1096
3	Lincoln Memorial	Jefferson Memorial	Jul	1047
4	Jefferson Dr & 14th St SW	Jefferson Dr & 14th St SW	Jul	1031
5	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Jul	982
6	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Apr	962
7	Lincoln Memorial	Jefferson Memorial	Aug	943
8	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Smithsonian-National Mall / Jefferson Dr & 12th St SW	Jun	916
9	Jefferson Dr & 14th St SW	Lincoln Memorial	Jun	850
10	Lincoln Memorial	Jefferson Memorial	Apr	837

... with 445,129 more rows

We Have Dock Station Coordinates

```
Simple feature collection with 508 features and 1 field
geometry type:  POINT
dimension:      XY
bbox:           xmin: -77.36842 ymin: 38.79756 xmax: -76.84326 ymax: 39.126
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
# A tibble: 508 x 2
   name                                        geometry
   <chr>                                     <POINT [°]>
 1 20th & E St NW                          (-77.045 38.8963)
 2 21st & I St NW                          (-77.047 38.9008)
 3 13th St & New York Ave NW              (-77.02982 38.90028)
 4 7th & F St NW / National Portrait Gallery  (-77.0222 38.8973)
 5 17th & K St NW / Farragut Square        (-77.03832 38.90206)
 6 3rd & M St NE                           (-77.00213 38.90547)
 7 Eads St & 15th St S                     (-77.05323 38.85897)
 8 18th & Eads St.                         (-77.05332 38.85725)
 9 Crystal Dr & 20th St S                  (-77.04923 38.85643)
10 Crystal Dr & 15th St S                  (-77.04959 38.86017)
# ... with 498 more rows
```

Similar to Node List

```
Simple feature collection with 508 features and 1 field
geometry type:  POINT
dimension:      XY
bbox:           xmin: -77.36842 ymin: 38.79756 xmax: -76.84326 ymax: 39.126
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
# A tibble: 508 x 2
   name                                  geometry
   <chr>                              <POINT [°]>
 1 20th & E St NW                  (-77.045 38.8963)
 2 21st & I St NW                  (-77.047 38.9008)
 3 13th St & New York Ave NW      (-77.02982 38.90028)
 4 7th & F St NW / National Portrait Gallery  (-77.0222 38.8973)
 5 17th & K St NW / Farragut Square (-77.03832 38.90206)
 6 3rd & M St NE                  (-77.00213 38.90547)
 7 Eads St & 15th St S            (-77.05323 38.85897)
 8 18th & Eads St.                (-77.05332 38.85725)
 9 Crystal Dr & 20th St S         (-77.04923 38.85643)
10 Crystal Dr & 15th St S         (-77.04959 38.86017)
# ... with 498 more rows
```

Coordinates are Similar to Node Attributes

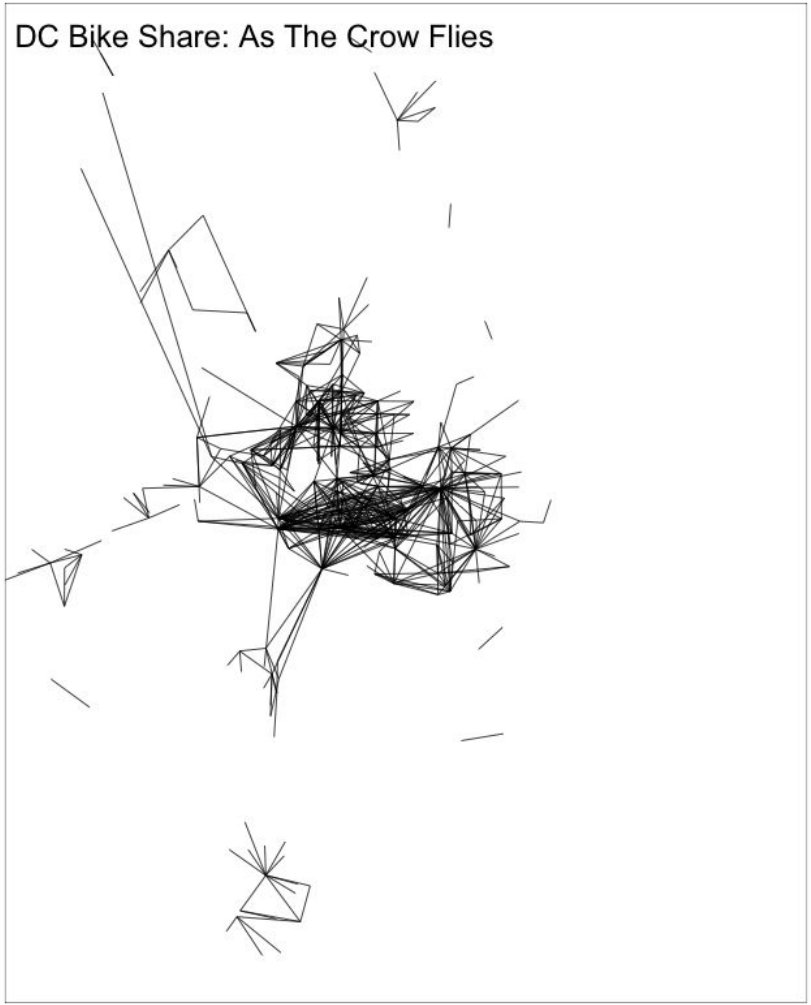
```
Simple feature collection with 508 features and 1 field
geometry type:  POINT
dimension:      XY
bbox:           xmin: -77.36842 ymin: 38.79756 xmax: -76.84326 ymax: 39.126
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
# A tibble: 508 x 2
   name                                    geometry
   <chr>                                <POINT [°]>
 1 20th & E St NW                    (-77.045 38.8963)
 2 21st & I St NW                    (-77.047 38.9008)
 3 13th St & New York Ave NW        (-77.02982 38.90028)
 4 7th & F St NW / National Portrait Gallery  (-77.0222 38.8973)
 5 17th & K St NW / Farragut Square (-77.03832 38.90206)
 6 3rd & M St NE                    (-77.00213 38.90547)
 7 Eads St & 15th St S              (-77.05323 38.85897)
 8 18th & Eads St.                  (-77.05332 38.85725)
 9 Crystal Dr & 20th St S           (-77.04923 38.85643)
10 Crystal Dr & 15th St S           (-77.04959 38.86017)
# ... with 498 more rows
```

Create a Spatial Line With stplanr::od2line()

```
Simple feature collection with 670 features and 2 fields
geometry type:  LINESTRING
dimension:      XY
bbox:           xmin: -77.35972 ymin: 38.79756 xmax: -76.97483 ymax: 39.126
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
First 10 features:
                                       start                                                  end                       geometry
1                          Jefferson Memorial                                     Lincoln Memorial LINESTRING (-77.03741 38.87...
2                 Jefferson Dr & 14th St SW                                     Lincoln Memorial LINESTRING (-77.03243 38.88...
3                            Lincoln Memorial Smithsonian-National Mall / Jefferson Dr & 12th St SW LINESTRING (-77.04943 38.88...
4              17th St & Independence Ave SW                                     Lincoln Memorial LINESTRING (-77.03833 38.88...
5                               6th & H St NE                       Columbus Circle / Union Station LINESTRING (-76.99835 38.89...
6  Henry Bacon Dr & Lincoln Memorial Circle NW                           Jefferson Dr & 14th St SW LINESTRING (-77.04938 38.89...
7                   Jefferson Dr & 14th St SW                                   Jefferson Memorial LINESTRING (-77.03243 38.88...
8                               8th & F St NE                       Columbus Circle / Union Station LINESTRING (-76.99475 38.89...
9                   Jefferson Dr & 14th St SW Smithsonian-National Mall / Jefferson Dr & 12th St SW LINESTRING (-77.03243 38.88...
10                     4th St & Madison Dr NW                           Jefferson Dr & 14th St SW LINESTRING (-77.01725 38.89...
```

LINESTRING Joins Dock Station Coordinates

```
Simple feature collection with 670 features and 2 fields
geometry type:  LINESTRING
dimension:      XY
bbox:           xmin: -77.35972 ymin: 38.79756 xmax: -76.97483 ymax: 39.126
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
First 10 features:
                                         start                                               end                       geometry
1                           Jefferson Memorial                                  Lincoln Memorial LINESTRING (-77.03741 38.87...
2                      Jefferson Dr & 14th St SW                                Lincoln Memorial LINESTRING (-77.03243 38.88...
3                             Lincoln Memorial Smithsonian-National Mall / Jefferson Dr & 12th St SW LINESTRING (-77.04943 38.88...
4                  17th St & Independence Ave SW                                Lincoln Memorial LINESTRING (-77.03833 38.88...
5                                 6th & H St NE                        Columbus Circle / Union Station LINESTRING (-76.99835 38.89...
6   Henry Bacon Dr & Lincoln Memorial Circle NW                        Jefferson Dr & 14th St SW LINESTRING (-77.04938 38.89...
7                      Jefferson Dr & 14th St SW                              Jefferson Memorial LINESTRING (-77.03243 38.88...
8                                 8th & F St NE                        Columbus Circle / Union Station LINESTRING (-76.99475 38.89...
9                      Jefferson Dr & 14th St SW Smithsonian-National Mall / Jefferson Dr & 12th St SW LINESTRING (-77.03243 38.88...
10                        4th St & Madison Dr NW                        Jefferson Dr & 14th St SW LINESTRING (-77.01725 38.89...
```


DC Bike Share: As The Crow Flies

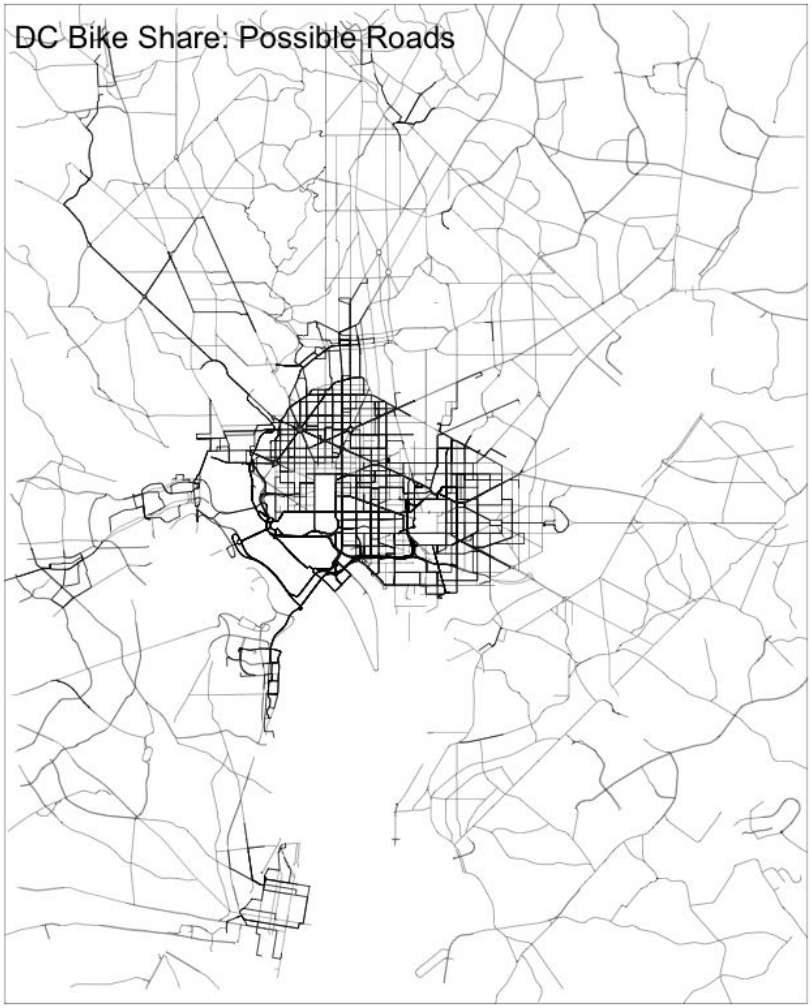
Guess Actual Routes With stplanr::line2route()

```
Simple feature collection with 7823 features and 8 fields (with 32 geometries empty)
geometry type:  MULTILINESTRING
dimension:      XY
bbox:           xmin: -77.12713 ymin: 38.79701 xmax: -76.9701 ymax: 38.9928
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
First 10 features:
   distance duration error id              start              end month   nn                       geometry
1    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Jul 1626 MULTILINESTRING ((-77.03741...
2    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Aug 1514 MULTILINESTRING ((-77.03741...
3    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Apr 1384 MULTILINESTRING ((-77.03741...
4    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Jun 1262 MULTILINESTRING ((-77.03741...
5    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Sep 1174 MULTILINESTRING ((-77.03741...
6    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   May 1122 MULTILINESTRING ((-77.03741...
7    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Oct 1044 MULTILINESTRING ((-77.03741...
8    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Mar  661 MULTILINESTRING ((-77.03741...
9    1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Nov  546 MULTILINESTRING ((-77.03741...
10   1655.7    274.7    NA  1 Jefferson Memorial Lincoln Memorial   Feb  486 MULTILINESTRING ((-77.03741...
> 
```


DC Bike Share: OSM Most Direct Routeś

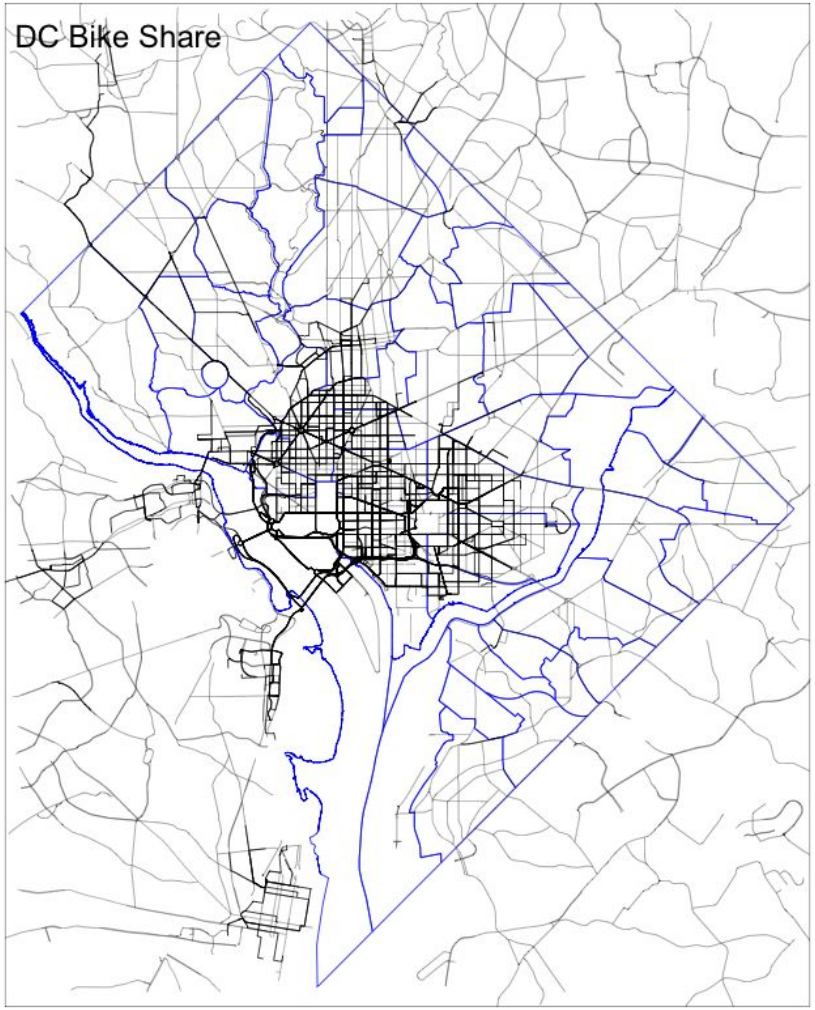
Get More Roads With osmdata::opq()

```
Simple feature collection with 67883 features and 1 field
geometry type:  LINESTRING
dimension:      XY
bbox:           xmin: -77.16399 ymin: 38.7555 xmax: -76.8774 ymax: 39.01832
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
First 10 features:
             highway                          geometry
4681060      service LINESTRING (-77.04818 38.83...
4681061      service LINESTRING (-77.05178 38.83...
4681062      service LINESTRING (-77.05148 38.83...
4681077 motorway_link LINESTRING (-77.10049 38.83...
4681080 motorway_link LINESTRING (-77.09205 38.83...
4681186  residential LINESTRING (-77.10826 38.93...
4725152    secondary LINESTRING (-76.93866 38.88...
5041550     motorway LINESTRING (-77.15598 38.88...
5041617     motorway LINESTRING (-77.07374 38.89...
5150071     cycleway LINESTRING (-77.11151 38.80...
```


DC Bike Share: Possible Roads

ShapeFiles Often Provide Useful Border Info

```
Simple feature collection with 46 features and 7 fields
geometry type:  POLYGON
dimension:      XY
bbox:           xmin: -77.1198 ymin: 38.79164 xmax: -76.90915 ymax: 38.99597
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
First 10 features:
   OBJECTID                WEB_URL       NAME                                                                NBH_NAMES Shape_Leng Shape_Area     TYPE                       geometry
1         1 http://planning.dc.gov/ Cluster 39                                Congress Heights, Bellevue, Washington Highlands  10711.668    4886463 Original POLYGON ((-76.99402 38.8452...
2         2 http://planning.dc.gov/ Cluster 38                                                    Douglas, Shipley Terrace   8229.486    2367958 Original POLYGON ((-76.97472 38.8528...
3         3 http://planning.dc.gov/ Cluster 36                              Woodland/Fort Stanton, Garfield Heights, Knox Hill   4746.344    1119573 Original POLYGON ((-76.96877 38.8606...
4         4 http://planning.dc.gov/ Cluster 27                                                    Near Southeast, Navy Yard   7286.969    1619167 Original POLYGON ((-76.98726 38.8771...
5         5 http://planning.dc.gov/ Cluster 32                                       River Terrace, Benning, Greenway, Dupont Park  11251.013    4286254 Original POLYGON ((-76.9376 38.88996...
6         6 http://planning.dc.gov/  Cluster 8 Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street   7343.049    2937442 Original POLYGON ((-77.00906 38.9056...
7         7 http://planning.dc.gov/  Cluster 5                                                    West End, Foggy Bottom, GWU   5884.746    1358616 Original POLYGON ((-77.05011 38.9089...
8         8 http://planning.dc.gov/ Cluster 30                                                 Mayfair, Hillbrook, Mahaning Heights   7599.066    2524927 Original POLYGON ((-76.94755 38.9059...
9         9 http://planning.dc.gov/ Cluster 31                  Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights   8570.353    3330854 Original POLYGON ((-76.91322 38.8897...
10       10 http://planning.dc.gov/  Cluster 7                                                                  Shaw, Logan Circle   5004.595    1463803 Original POLYGON ((-77.03195 38.9141...
```


DC Bike Share

Spatial Features Can Contain Data

```
Simple feature collection with 2363 features and 11 fields
geometry type:  MULTILINESTRING
dimension:      XY
bbox:           xmin: -77.10797 ymin: 38.80351 xmax: -76.9701 ymax: 38.97801
epsg (SRID):    4326
proj4string:    +proj=longlat +datum=WGS84 +no_defs
First 10 features:
   distance duration error id                                      start                                 end month  nn n.overlaps origins agg
1    5418.2    558.3  <NA>  2                     Jefferson Dr & 14th St SW                     Lincoln Memorial   Jan 234          1       1 175
2    1307.6    263.9  <NA>  5                                6th & H St NE       Columbus Circle / Union Station   Jan 213          1       2 234
3    1282.7    280.4  <NA>  8                                8th & F St NE       Columbus Circle / Union Station   Jan 570          1       3 104
4     653.0     76.7  <NA> 15                  Adams Mill & Columbia Rd NW            Calvert St & Woodley Pl NW   Jan 396          1       4 107
5    1515.0    256.5  <NA> 17 Eastern Market Metro / Pennsylvania Ave & 7th St SE Lincoln Park / 13th & East Capitol St NE   Jan 342          1       5 213
6    1982.2    305.7  <NA> 19                               13th & H St NE       Columbus Circle / Union Station   Jan 205          1       6 118
7    1663.7    302.4  <NA> 21                  Columbus Circle / Union Station              Maryland Ave & E St NE   Jan 317          1       7 117
8    1809.0    306.6  <NA> 24                               11th & H St NE       Columbus Circle / Union Station   Jan 350          1       8 570
9    2212.7    354.3  <NA> 33                               13th & D St NE       Columbus Circle / Union Station   Jan 282          1       9  70
10   1256.2    286.4  <NA> 34                               15th & P St NW  Massachusetts Ave & Dupont Circle NW   Jan 337          1      10  85
                        geometry
1  MULTILINESTRING ((-77.03236...
2  MULTILINESTRING ((-76.99844...
3  MULTILINESTRING ((-76.99475...
4  MULTILINESTRING ((-77.04275...
5  MULTILINESTRING ((-76.99527...
6  MULTILINESTRING ((-76.98831...
7  MULTILINESTRING ((-77.00058...
8  MULTILINESTRING ((-76.99153...
9  MULTILINESTRING ((-76.98763...
10 MULTILINESTRING ((-77.03456...
```

Jan

DC Bike Share Routes Over Time

Bikes

- < 450
- < 900
- <1350k
- >1350

DC Bike Share Routes Over Time

Bikes

- < 450
- < 900
- <1350k
- >1350

Resources

Geocomputation with R

Robin Lovelace, Jakub Nowosad, Jannes Muenchow

2018-09-27

https://geocompr.robinlovelace.net/
@robinlovelace

Also check out: **@bikesRdata**

Hands on Intro