

Two neat **tidymetrics** functions to make calculating metrics (and life) more fun

Kaelen L. Medeiros

@kaelen_medeiros

Slides @ bit.ly/klm-rstudio-2020

tidymetrics

- Available on [GitHub](#)
- “Dimensional modeling done the tidy way” - the documentation
- Extensive uses for calculating metrics

cross_by_periods()

“Cross by any set of calendar periods (like day or week), rolling windows, or recent intervals (like “4 Weeks”, or “8 Weeks”).”

MUST have a column of data called “date” with JUST dates in it.

For example...

```
> questions %>% select(owner_user_id, date, tag, everything())
```


```
# A tibble: 76,616 x 8
```

	owner_user_id	date	tag	id	creation_date	score	title	body
	<dbl>	<date>	<chr>	<dbl>	<dtm>	<dbl>	<chr>	<chr>
1	15842	2008-09-18	statist...	95007	2008-09-18 17:59:19	56	Explain the quantil...	"<p>I've been mystified by the ...
2	1941213	2008-11-01	statist...	255697	2008-11-01 15:48:30	4	Is there an R packa...	"<p>I'm looking for a an <code>...
3	67405	2009-02-18	statist...	560329	2009-02-18 09:08:38	2	Sort the X axis in ...	"<p>I have binned data that loo...
4	69117	2009-02-27	None	596819	2009-02-27 21:49:17	11	What is the best wa...	"<p>I have 12 data frames to wo...
5	69117	2009-02-27	None	596976	2009-02-27 22:45:59	0	What is the Y funct...	"<p>A friend of mine asked me i...
6	67405	2009-03-18	statist...	657440	2009-03-18 08:59:04	7	Cumulative Plot wit...	"<p>I have data that looks like...
7	20895	2009-04-10	statist...	736514	2009-04-10 02:18:38	41	R Random Forests Va...	"<p>I am trying to use the rand...
8	85950	2009-04-10	statist...	736541	2009-04-10 02:32:21	12	Plots without title...	"<p>In <a href=\"http://www.r-p...
9	14744	2009-04-13	None	743812	2009-04-13 12:52:51	125	Calculating moving ...	"<p>I'm trying to use R to calc...
10	85950	2009-04-15	statist...	750703	2009-04-15 07:55:21	28	"Suppressing \"null...	"<p>I have a number of bash scr...

```
# ... with 76,606 more rows
```

of questions in weeks, months

```
```{r}
questions %>%
 cross_by_periods(c("week", "month")) %>%
 summarize(nb_questions = n()) %>%
 arrange(desc(date))
```
```




A tibble: 544 x 3
Groups: period [2]

| | period | date | nb_questions |
|----|--------|------------|--------------|
| | <chr> | <date> | <int> |
| 1 | week | 2017-09-18 | 504 |
| 2 | week | 2017-09-11 | 493 |
| 3 | week | 2017-09-04 | 468 |
| 4 | month | 2017-09-01 | <u>1588</u> |
| 5 | week | 2017-08-28 | 448 |
| 6 | week | 2017-08-21 | 484 |
| 7 | week | 2017-08-14 | 427 |
| 8 | week | 2017-08-07 | 521 |
| 9 | month | 2017-08-01 | <u>2181</u> |
| 10 | week | 2017-07-31 | 501 |

... with 534 more rows

Rolling 28-day # of questions

```
```{r}
questions %>%
 cross_by_periods(windows = c(28)) %>%
 summarize(nb_questions = n()) %>%
 arrange(desc(date))
```
```



```
# A tibble: 3,800 x 3
# Groups:   period [4]
  period      date      nb_questions
  <chr>      <date>         <int>
1 rolling_28d 2017-10-21         48
2 rolling_28d 2017-10-20         83
3 rolling_28d 2017-10-19        155
4 rolling_28d 2017-10-18        246
5 rolling_28d 2017-10-17        327
6 rolling_28d 2017-10-16        410
7 rolling_28d 2017-10-15        504
8 rolling_28d 2017-10-14        544
9 rolling_28d 2017-10-13        577
10 rolling_28d 2017-10-12        660
# ... with 3,790 more rows
```

**Great for rolling
averages!**

cross_by_dimensions()

“...acts as an extended group_by that allows complete summaries across each individual dimension and possible combinations.”

questions per tag

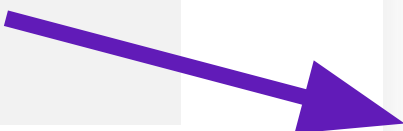
```
```{r}
questions %>%
 cross_by_dimensions(tag) %>%
 summarize(nb_questions = n())
```
```

Note the new “All” dimension!!

```
# A tibble: 10 x 2
  tag          nb_questions
<chr>          <int>
1 All          76616
2 data.table   5632
3 dplyr        5873
4 ggplot2     17210
5 None        30894
6 package     1079
7 python      2132
8 regex       3079
9 shiny       8064
10 statistics  2653
```


More powerful with two!

```
```{r}
questions %>%
 cross_by_periods("week") %>%
 cross_by_dimensions(tag) %>%
 summarize(nb_questions = n()) %>%
 arrange(desc(date)) %>%
 head(20)
```
```



```
# A tibble: 20 x 4
# Groups:   period, date [2]
   period date      tag      nb_questions
   <chr>   <date>   <chr>         <int>
1 week   2017-09-18 All           504
2 week   2017-09-18 data.table     26
3 week   2017-09-18 dplyr          58
4 week   2017-09-18 ggplot2       118
5 week   2017-09-18 None          182
6 week   2017-09-18 package         8
7 week   2017-09-18 python         12
8 week   2017-09-18 regex          18
9 week   2017-09-18 shiny          62
10 week  2017-09-18 statistics     20
11 week  2017-09-11 All           493
12 week  2017-09-11 data.table     28
13 week  2017-09-11 dplyr          84
14 week  2017-09-11 ggplot2       105
15 week  2017-09-11 None          152
16 week  2017-09-11 package         2
17 week  2017-09-11 python         11
18 week  2017-09-11 regex          19
19 week  2017-09-11 shiny          87
20 week  2017-09-11 statistics      5
```

Kaelen L. Medeiros

Hire me in NYC or remote!!!

@kaelen_medeiros

kaelenmedeiros@gmail.com

klmedeiros.com

Slides @ bit.ly/klm-rstudio-2020

