

SPRINTING IN JUPYTER THROUGH R TRACK RESULTS

RICK PACK (LABCORP DATA SCIENTIST)



https://www.carolinagodiva.org/index.php?page=summer-track-5-30-2018



**carolina
Godiva**
TRACK • CLUB
RUNNING SINCE 1975

Join CGTC or Renew Your Membership

Home Club Info Join Us Weekly Runs Racing Photos Newsletter Contact

Summer Track 5-30-2018

Mile Run				
Jan Alexander	M	39	5:06	
Rob Benjamin	M	49	5:15	
Hayes Weatherman	M	13	5:35	
Leif Rasmussen	M	15	5:37	
Karna Morey	M	17	5:38	
Ary Bush	M	38	5:49	
Lia Weiner	F	29	5:50	
Jackson Steffens	M	12	5:53	
Kevin Nickodem	M	61	5:53	

Enter search...

Submit

Find us on Facebook

Umstead Trail Marathon

Our Latest Newsletter

Join CGTC or Renew Your Membership



Louise Guardino, CGTC President Bill Harris, and honorable starter Ethan Caldwell

```

In [56]: # World Masters Athletics (WMA) Age-graded track tables can be found on
# Howard Grubb's website:
# http://www.howardgrubb.co.uk/athletics/data/wavacalc06.xls

library(rvest)
library(readr)
library(stringr)
library(dplyr)
library(tidyr)
library(readxl)
library(lubridate)
library(patchwork)
library(ggplot2)

setwd('C:/Users/Packr1/Documents/Personal/Track')
WMA_Women <- read_xls('wavacalc06.xls', sheet = 1) %>%
  select(2:ncol()) %>%
  mutate(dist_m = 1000 * `dist(km)`) %>%
  mutate(Sex = 'F') %>%
  dplyr::filter(dist_m > 0 &
    !grepl('Road', Event))
print(nrow(WMA_Women) - nrow(WMA_Women %>% distinct(dist_m)))
message("Expecting 0 so no duplicate distance")
WMA_Men <- read_xls('wavacalc06.xls', sheet = 2) %>%
  select(2:ncol()) %>%
  mutate(dist_m = 1000 * `dist(km)`) %>%
  mutate(Sex = 'M') %>%
  dplyr::filter(dist_m > 0 &
    !grepl('Road', Event))
print(nrow(WMA_Men) - nrow(WMA_Men %>% distinct(dist_m)))
message("Expecting 0 so no duplicate distance")

track_file <- "C:/Users/Packr1/Documents/Personal/Track/track_res.csv"
track_res <- read_csv(track_file)
pics_dir <- "C:/Users/Packr1/Documents/Personal/Track/Godiva/Jupyter_Pics/"

[1] 0
Expecting 0 so no duplicate distance

[1] 0
Expecting 0 so no duplicate distance
Parsed with column specification:
cols(
  Name = col_character(),
  Sex = col_character(),
  Age = col_integer(),
  Time = col_character(),
  ct_evt = col_integer(),
  Event = col_character(),
  mins = col_integer(),

```

```

library(rvest)
library(readr)
library(stringr)
library(dplyr)
library(tidyr)
library(readxl)
library(lubridate)
library(patchwork)
library(ggplot2)

```

CTRL+ENTER

R: EXECUTES A SINGLE LINE

JUPYTER NOTEBOOK: RUNS AN ENTIRE CELL

Use `installed.packages()` TO SEE WHERE
PACKAGES STORED

DieselAnalytics / TriPASS

Unwatch 2 Unstar 1 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Branch: master TriPASS / RVizInJupyterLab.ipynb Find file Copy path

DieselAnalytics Add files via upload b60de0f 20 days ago

1 contributor

995 lines (994 sloc) 35.2 KB

Raw Blame History

Creating Custom R Visualizations in Jupyter Lab using ggplot2

Prerequisites

- Install Jupyter Lab
 - Install Anaconda. Get installer [here](#).
 - Install Jupyter Lab
 - Open up the Anaconda Command Prompt
 - Enter the following code snippet:

```
conda install -c conda-forge jupyterlab
```
- Instructions of how to install the R Kernel in Jupyter Notebook can be found in this [YouTube](#) video.
- [List](#) of available kernels for Jupyter Notebooks.
- [Link](#) to documentation for iPython magics.
- [Link](#) to documentation for R magics.
- [Link](#) to useful Jupyter Notebook tips.
- [Link](#) to Jupyter Notebook keyboard shortcuts.
- [Link](#) to a *Data Framed* podcast about Jupyter Lab and the direction of the Jupyter project.

GETTING STARTED WITH JUPYTER
NOTEBOOKS (R)

Ryan Wade's TriPASS presentation

```
In [57]: track_res %>% group_by(Name, Sex, Event, Date_Meet) %>%  
          summarise(ct = n()) %>% dplyr::filter(ct > 1) %>% ungroup()
```

Name	Sex	Event	Date_Meet	ct
Aimee Bazin	F	100m Dash	2014-07-23	2
Aimee Bazin	F	200m	2013-07-10	2
Alex Andre	M	200m	2013-07-10	2
Alex Emmons	M	Mile Run	2016-07-13	2
Alex Taylor	M	200m	2009-07-15	2
Aline Lloyd	F	200m	2013-07-10	2
Barbara Hindenach	F	100m Dash	2014-07-23	2
Barbara Hindenach	F	200m	2009-07-15	2
Barbara Hindenach	F	200m	2013-07-10	2
Barbara Hindenach	F	800 Meter Run	2016-06-29	2
Barbara Hindenach	F	800m	2013-05-29	2

USING DPLYR TO IDENTIFY
EXTRA RECORDS

Expected one record per Name, Sex,
Event, and Date

```
In [58]: track_res %>% dplyr::filter(str_trim(Name) == 'Aimee Bazin' &
                                     str_trim(toupper(Event)) == '100M DASH' &
                                     Date_Meet == '2014-07-23')
```

Name	Sex	Age	Time	ct_evt	Event	mins	secs	track_time	temp_dist	dist_m	Place	Max_Hamlyn_pts	Date_Meet	agegrp
Aimee Bazin	F	31	18.1	2000	100m Dash	0	18.1	18.1	100m	NA	6	0	2014-07-23	Age 30 - 34
Aimee Bazin	F	31	11:14	2000	100m Dash	11	14	674.0	100m	NA	21	0	2014-07-23	Age 30 - 34

USING DPLYR TO IDENTIFY
EXTRA RECORDS

Expected one record per Name, Sex,
Event, and Date

1500m Racewalk

Ray Dooley	M	61	8:38
Roxanne Springer	F	50	8:53
Barbara Hindenach	F	63	9:49
Jason Figge	M	41	10:11
Ben Barker	M	44	10:18
Brianna Honea	F	35	10:23
Tim O'Brien	M	62	10:43
Pete Gilligan	M	62	10:49
Lena Hollmann	F	63	10:50
William Schmitz	M	42	10:50
Courtney Tuszynski	F	35	10:56
Aimee Bazin	F	31	11:14

100m Dash

Web-scraping code
expected bold event
header

Jermaine Wade	M	30	11.4
Asad Ahmad	M	26	12.2
Kwame Alston	M	24	12.2
Mike Carpenter	M	46	12.2
Rick Pack	M	33	12.4
Amy Carpenter	F	15	12.7

USING DPLYR TO IDENTIFY
EXTRA RECORDS

Expected one record per Name, Sex,
Event, and Date


```
In [60]: # Select the fastest time whenever there are duplicates per name, event, date
track_res <- track_res %>%
  group_by(Name, Event, Date_Meet) %>%
  slice(which.min(track_time)) %>% ungroup()
track_res %>% group_by(Name, Sex, Event, Date_Meet) %>%
  summarise(ct = n()) %>% dplyr::filter(ct > 1)
```

Name	Sex	Event	Date_Meet	ct
------	-----	-------	-----------	----

```
In [61]: track_res %>% dplyr::filter(str_trim(Name) == 'Aimee Bazin' &
  str_trim(toupper(Event)) == '100M DASH' &
  Date_Meet == '2014-07-23')
```

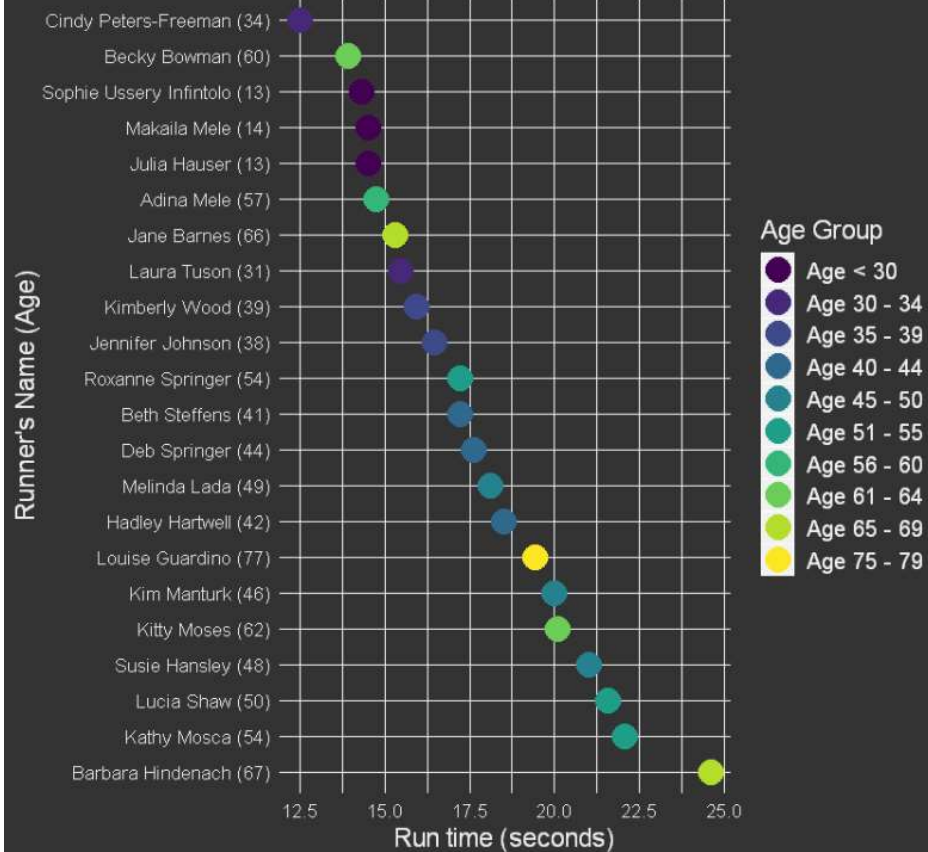
Name	Sex	Age	Time	ct_evt	Event	mins	secs	track_time	temp_dist	dist_m	Place	Max_Hamlyn_pts	Date_Meet	agegrp
Aimee Bazin	F	31	18.1	2000	100m Dash	0	18.1	18.1	100m	NA	6	0	2014-07-23	Age 30 - 34

USING JUPYTER TO TEACH #RSTATS

Ask mentees what number to use in the inequality

Carolina Godiva Track Club Summer Track Series (2018)

100 Meter: dot plot [top 3 Males per age group]



Becky Bowman leading a pack of courageous 100 meter USATF competitors

Carolina Godiva Track Club Summer Track Series (2018)

100 Meter: dot plot [top 3 per age group]

Runners (Name (Age)) listed on the y-axis:

- Rick Pack (37)
- Rayvon Moore (25)
- Dante Freeman (33)
- Leon Bullard (44)
- Gerald Mitchell (55)
- Eric Cabbell, Jr (23)
- Cindy Peters-Freeman (34)
- Leif Rasmussen (15)
- Elias Williamson (16)
- Kevin McCabe (37)
- Andrew McIver (36)
- Becky Bowman (60)
- Keith Volmar (47)
- Chris Sarsony (50)
- Andy Snipes (32)
- Ben Duncan (52)
- Adina Mele (57)
- Jane Barnes (66)
- Patrick Gale (44)
- Derek Rodriguez (50)
- Jesper Rasmussen (44)
- Stephen Fraser (48)
- Mike Nobles (55)
- Bobbie Perry (72)
- Kevin Nickodem (61)
- Gordon Keeler (63)
- Heiko Rath (48)
- Mike Valle (77)
- Louise Guardino (77)
- Ken Larsen (70)
- Tom Griffin (72)
- Tom Kirby (65)
- Barbara Hindenach (67)
- Richard Wolfe (80)

Age Group Legend:

- Age < 30
- Age 30 - 34
- Age 35 - 39
- Age 40 - 44
- Age 45 - 50
- Age 51 - 55
- Age 56 - 60
- Age 61 - 64
- Age 65 - 69
- Age 70 - 74
- Age 75 - 79
- Age 80 - 85

Run time (seconds) on the x-axis.

100 Meter: dot plot [top 3 per age group]

Runner's Name (Age)

Age Group



Age < 30
Age 30 - 34
Age 35 - 39
Age 40 - 44
Age 45 - 50
Age 51 - 55
Age 56 - 60
Age 61 - 64
Age 65 - 69
Age 70 - 74
Age 75 - 79
Age 80 - 85

Run time (seconds)

SPRINTING IN JUPYTER THROUGH R TRACK RESULTS

RICK PACK (LABCORP DATA SCIENTIST)



TEACHING R WITH JUPYTER NOTEBOOKS
ASKING DIVERSE POPULATIONS TO DANCE

Rick Pack
@rick_pack2 on Twitter
<http://rickpackblog.wordpress.com>