## Panel Shows and Swimmers

I really like are panel shows. We don't really have them in the US, but they're common in Britain, and available online. Generally speaking a panel show is a type of television program where a host and a number of panelists undertake a game or conversation in an entertaining fashion. Panelists are usually stand up comedians but sometimes other notables, like athletes, participate as well. Olympic gold medalist Rebecca Adlington was a panelist on 8 Out of Ten Cats ("a show about statistics" as the tag line goes) after the London Games.

*Rebecca Adlington joins Comedians Jon Richardson and Romesh Ranganathan*

After the Rio games gold medalist and Paralympian Ellie Simmonds was on as well and demonstrated her skill at a "cereal box game". When it comes to having swimmers on as guests though no show does better than the Last Leg. They've had lots of swimmers. Liz Johnson, Sasha Kindred, Jeanette Chippington, the aforementioned Ellie Simmonds, and plenty more.

*Ellie Simmonds on the Last Leg*

I watch that show all the time and it brings me a lot of joy. Host Adam Hills frequently challenges people to do better, often specifically advocating for improved access for people with disabilities.

So, as you may have guessed from the post title, we here at Swimming + Data Science are attempting to meet Hillsy's challenge by better addressing para athletics within `SwimmeR`. As of v0.8.0 `SwimmeR` now handles para swimming codes (S4, SM10 etc.).

## Setup

First download the new version from CRAN.

```
install.packages("SwimmeR")
```

Then load the package and some others that we'll also need.

```
library(SwimmeR)
library(flextable)
library(dplyr)

flextable_style <- function(x) {
  x %>%
    flextable() %>%
    bold(part = "header") %>% # bolds header
    bg(bg = "#D3D3D3", part = "header") %>%  # puts gray background behind the
header row
    autofit()
}
```

## Para Codes

We can take a look at results from the 2020 Jimi Flowers meet, the most recent meet results hosted on the U.S. Paralympic Swimming results repository.

```
file <- "https://raw.githubusercontent.com/gpilgrim2670/Pilgrim_Data/master/2020_Jimi_Flowers_
Results_PDF.pdf"

df <- swim_parse(read_results(file))

df %>%
  head(10) %>%
  flextable_style()
```

| Place | Name | Age | Para | Team | Prelims_Time | Finals_Time | DQ | Exhibition | Event |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Smith, Leanne | 31 | S3 | US Paralympics Resident Team-CO- | 44.28 | 42.96 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S3 |
| 2 | Ramirez Martinez, Fabiola | 29 | S3 | Jalisco- | 1:13.10 | 1:12.17 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S3 |
| 1 | Locatelli, Wendi | 37 | S5 | Unattached- | 49.00 | 47.73 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S5 |
| 2 | Hernandez Torres, Karina Ama | 25 | S5 | Jalisco- | 53.10 | 54.00 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S5 |
| 3 | Pareé , Cleé mence | 17 | S5 | Unattached-CAN | 54.43 | 57.09 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S5 |

| Place | Name | Age | Para | Team | Prelims_Time | Finals_Time | DQ | Exhibition | Event |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Lomeli Santos, Nancy Nayely | 23 | S6 | Jalisco- | 40.35 | 41.37 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S6 |
| 2 | Bravo Gonzalez, Karla France | 21 | S6 | Jalisco- | 41.30 | 43.03 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S6 |
| 1 | Coan, McKenzie | 23 | S7 | Cumming Waves Swim Team-GA- | 32.52 | 33.46 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S7 |
| 2 | Weggemann, Mallory | 30 | S7 | Unattached- | 33.00 | 34.43 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S7 |
| 3 | Gaffney, Julia | 19 | S7 | US Paralympics Resident Team-CO- | 34.30 | 35.39 | 0 | 0 | Women 50 LC Meter Freestyle Multi-Class S7 |

Note the addition of a new column, `Para`, containing paralympic classification codes parsed from the result. It's not a big change, but those codes are literally the only difference between para and non-para swimming results.

## Names

We've discussed names here before, specifically the "records matching" problem. It's probably the trickiest problem in dealing with swimming results, which is the aim of `SwimmeR`. There aren't any perfect solutions. Still, we're plugging away. Version 0.9.0 contains our latest contribution to the issue.

Names in swimming results aren't presented in a consistent format. Sometimes they're done as Firstname Lastname (Lilly King), sometimes as Lastname, Firstname (King, Lilly). This is simple enough for athletes with only one first or last name, but imagine a swimmer named Kara Lynn Joyce. There's no way to tell just based on the name itself if she should be Lynn Joyce, Kara or Joyce, Kara Lynn. What this means is that while there's more information encoded in Lastname, Firstname (because the comma differentiates between Lastname, however long, and Firstname, however long) the default format must be Firstname Lastname. It's simply not possible to rigorously convert Firstname Lastname to Lastname, Firstname based on the information available.

Enter the `name_reorder` function. `name_reorder` works on lists or whole data frames.

### Lists

Passing a list to `name_reorder` is simpler and more general than passing a data frame, just outputting a list with the names reordered to "Firstname Lastname".

```
name_examples_list <- c("Kara Lynn Joyce", "Joyce, Kara Lynn", "de Bruijn, Inge",
"Inge de Bruijn", NA)

name_examples_list %>%
  name_reorder()
## [1] "Kara Lynn Joyce" "Kara Lynn Joyce" "Inge de Bruijn"  "Inge de Bruijn"
## [5] NA
```

Since columns in a data frame are really just lists this also works with dplyr functions like `mutate`.

```
name_examples_dplyr <- data.frame(Athlete = c("Kara Lynn Joyce", "Joyce, Kara
Lynn", "de Bruijn, Inge", "Inge de Bruijn", NA))

name_examples_dplyr %>%
  mutate(Name = name_reorder(Athlete)) %>%
  flextable_style()
```

| Athlete | Name |
|---|---|
| Kara Lynn Joyce | Kara Lynn Joyce |
| Joyce, Kara Lynn | Kara Lynn Joyce |
| de Bruijn, Inge | Inge de Bruijn |
| Inge de Bruijn | Inge de Bruijn |

**Data Frames**

In contrast to usage with lists using `name_reorder` with entire data frames has a very `SwimmeR`-centric flavor. When given a data frame `name_reoder` converts all names, in a column called “Name” (to match the output of `swim_parse`) to Firstname Lastname format. By default the output is a data frame with one extra column, called `Name_Reorder`.

```
name_examples_df <- data.frame(Name = c("Kara Lynn Joyce", "Joyce, Kara Lynn",
"de Bruijn, Inge", "Inge de Bruijn", NA))

name_examples_df %>%
  name_reorder() %>%
  relocate(Name) %>% # want Name column first for presentation
  flextable_style()
```

| Name | Name_Reorder |
|---|---|
| Kara Lynn Joyce | Kara Lynn Joyce |
| Joyce, Kara Lynn | Kara Lynn Joyce |
| de Bruijn, Inge | Inge de Bruijn |
| Inge de Bruijn | Inge de Bruijn |

Setting the optional argument `verbose = TRUE` will add additional columns `First_Name` and `Last_Name` if extracting them is possible. This is perhaps helpful to people like me with an interest in names.

```
name_examples_df %>%
  name_reorder(verbose = TRUE) %>%
  relocate(Name) %>% # want Name column first for presentation
```

```
  flextable_style()
```

| Name | Name_Reorder | First_Name | Last_Name |
|---|---|---|---|
| Kara Lynn Joyce | Kara Lynn Joyce | | |
| Joyce, Kara Lynn | Kara Lynn Joyce | Kara Lynn | Joyce |
| de Bruijn, Inge | Inge de Bruijn | Inge | de Bruijn |
| Inge de Bruijn | Inge de Bruijn | | |

With `name_reorder` one can insure that a data set comprised of results from several meets will have all names in a consistent format. This is the first step in series of several planned additions to `SwimmeR` aimed at addressing name-related issues.

## Split Distances

We've discussed splits before, in conjunction with the `splits` and `splits_length` arguments to `swim_parse`. The idea is simple: setting `splits = TRUE` causes splits to be collected in columns, with the column names based on `splits_length`. There's a problem though when some events in a set of results have different split lengths than others. Consider the 2021 Women's NCAA DI championships.

```
file <- "https://s3.amazonaws.com/sidearm.sites/gopack.com/documents/2021/3/20/2021_DI_Women_Final_Results.pdf"

DI_W_2021 <- swim_parse(read_results(file), splits = TRUE, split_length = 50)
```

Most of the events are split by 50, except for the 50 Yard Freestyle and 200 Yard Freestyle Relay. They're split by 25, but the column names don't reflect that.

```
DI_W_2021 %>%
  filter(Event %in% c("Women 50 Yard Freestyle", "Women 200 Yard Freestyle Relay", "Women 200 Yard Freestyle")) %>%
  select(Place, Team, Event, Finals_Time, Split_50:Split_400) %>%
  group_by(Event) %>%
  slice_head() %>%
  flextable_style()
```

| Place | Team | Event | Finals_Time | Split_50 | Split_100 | Split_150 | Split_200 | Split_250 | Split_300 | Split_350 | Split_400 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Virginia | Women 200 Yard Freestyle | 1:42.35 | 24.13 | 25.60 | 25.91 | 26.71 | | | | |
| 1 | California | Women 200 Yard Freestyle Relay | 1:25.78 | 10.82 | 22.09 | 10.02 | 21.23 | 10.18 | 21.24 | 10.05 | 21.22 |

| Place | Team | Event | Finals_Time | Split_50 | Split_100 | Split_150 | Split_200 | Split_250 | Split_300 | Split_350 | Split_400 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Virginia | Women 50 Yard Freestyle | 21.13 | 10.33 | 10.80 | | | | | | |

We can fix this issue with the new `correct_split_distance` function. It will rename columns in the indicated `events` based on a `new_split_length`. I recognized too late that this function should really be called `correct_split_length` and have *ahem* corrected this oversight via an alias in the latest dev version of SwimmeR.

```
DI_W_2021 %>%
  correct_split_distance(
    new_split_length = 25,
    events = c("Women 50 Yard Freestyle", "Women 200 Yard Freestyle Relay")
  ) %>%
  filter(
    Event %in% c(
      "Women 50 Yard Freestyle",
      "Women 200 Yard Freestyle Relay",
      "Women 200 Yard Freestyle"
    )
  ) %>%
  group_by(Event) %>%
  select(
    Place,
    Team,
    Event,
    Finals_Time,
    Split_25,
    Split_50,
    Split_75,
    Split_100,
    Split_125,
    Split_150,
    Split_175,
    Split_200
  ) %>%
  slice_head() %>%
  flextable_style()
```

| Place | Team | Event | Finals_Time | Split_25 | Split_50 | Split_75 | Split_100 | Split_125 | Split_150 | Split_175 | Split_200 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | California | Women 200 Yard Freestyle Relay | 1:25.78 | 10.82 | 22.09 | 10.02 | 21.23 | 10.18 | 21.24 | 10.05 | 21.22 |
| 1 | Virginia | Women 50 Yard Freestyle | 21.13 | 10.33 | 10.80 | | | | | | |
| 1 | Virginia | Women 200 Yard | 1:42.35 | | 24.13 | | 25.60 | | 25.91 | | 26.71 |

| Place | Team | Event | Finals_Time | Split_25 | Split_50 | Split_75 | Split_100 | Split_125 | Split_150 | Split_175 | Split_200 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Freestyle | | | | | | | | | |