

Compare 2 Data Frames in R

Let's create the new data frames:

```
library(waldo)
```

```
df1<-data.frame(X=c(1,2,3), Y=c("a","b","c"), A=c(3,4,5))
df2<-data.frame(X=c(1,2,3,4), Y=c("A","b","c","d"),
Z=c("k","l","m","n"), A=c("3","4","5","6"))
```

The df1:

```
  X Y A
1 1 a 3
2 2 b 4
3 3 c 5
```

And the df2:

```
  X Y Z A
1 1 A k 3
2 2 b l 4
3 3 c m 5
4 4 d n 6
```

Let's compare them:

```
waldo::compare(df1,df2)
```

And we get:

```
> waldo::compare(df1,df2)
`old` is length 3
`new` is length 4

`names(old)` : "X" "Y"      "A"
`names(new)` : "X" "Y" "Z"  "A"

`attr(old, 'row.names')` : 1 2 3
`attr(new, 'row.names')` : 1 2 3 4

`old$X` : 1 2 3
`new$X` : 1 2 3 4

`old$Y` : "a" "b" "c"
`new$Y` : "A" "b" "c" "d"

`old$A` is a double vector (3, 4, 5)
`new$A` is a character vector ('3', '4', '5', '6')

`old$Z` is absent
`new$Z` is a character vector ('k', 'l', 'm', 'n')
```

As we can see it captures all the differences and the output is in a friendly format of different colors depending on the difference. More particularly, it captures that:

- The df1 has 3 rows where df2 has 2
- The df2 has an extra column name and it shows where is the difference in the order.
- It shows the rows names and the differences
- It compares each column, X, Y, A, Z
- It detected the difference in the data type of column A where in the first is double and in the second is character
- It returns that there is a new column called Z.