

# Tidyverse

## What is the “tidyverse”?

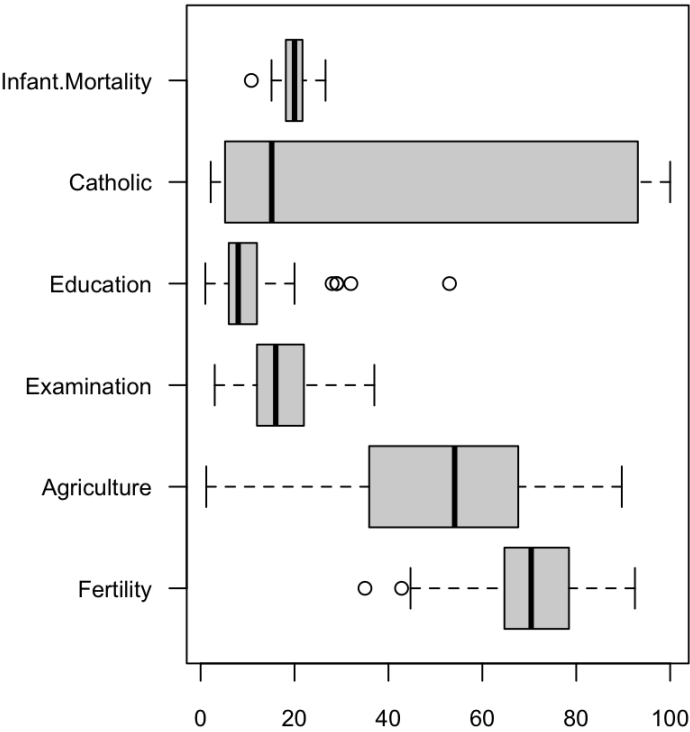
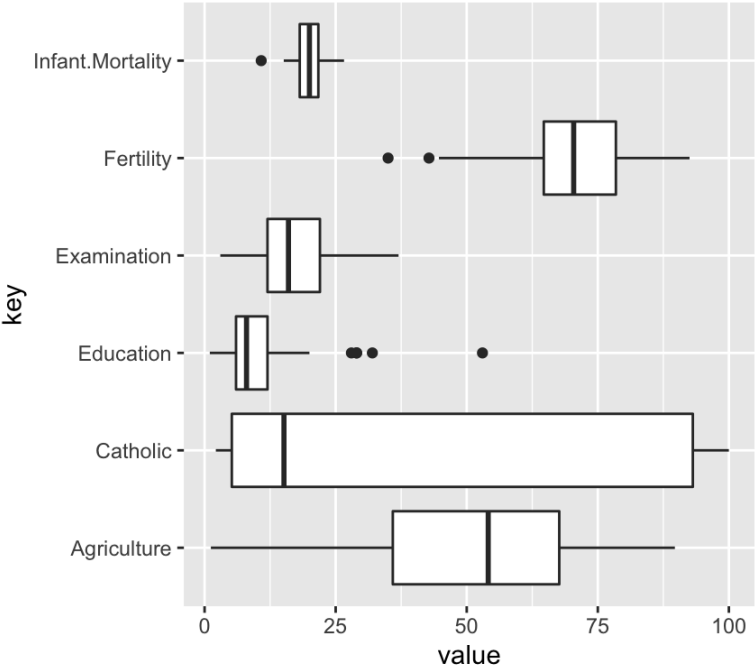
- “an opinionated collection of R packages designed for data science. All packages share an underlying philosophy and common APIs.”
- formerly referred to as the “hadleyverse” for Hadley Wickham
- packages are strongly associated with RStudio, but not exclusively

## Core tidyverse packages



- ggplot2
- dplyr
- tidyr
- readr
- purrr
- tibble
- stringr
- forcats

# 10 differences between the tidyverse and base R



# 1. Base R

Base:

```
barplot(1:5, horiz = TRUE)  
boxplot(1:5, horizontal = TRUE)
```

# 1. Tidyverse is ... more consistent

Tidyverse:

```
ggplot(...) + geom_bar() + coord_flip()  
ggplot(...) + geom_boxplot() + coord_flip()
```

## 2. Base R

Base:

```
df <- data.frame(x = 1:4, y = 1:2)
```

## 2. Base R

Base:

```
df <- data.frame(x = 1:4, y = 1:2)
```

```
df
```

```
##      x y  
## 1 1 1  
## 2 2 2  
## 3 3 1  
## 4 4 2
```



## 2. Tidyverse ... fails faster

Tidyverse:

```
df <- tibble(x = 1:4, y = 1:2)
```

## 2. Tidyverse ... fails faster

Tidyverse:

```
library(tibble)
df <- tibble(x = 1:4, y = 1:2)
```

```
## Error: Tibble columns must have compatible sizes.
## * Size 4: Existing data.
## * Size 2: Column `y`.
## i Only values of size one are recycled.
```

### 3. Base R

Base R used to prefer factors

```
df <- read.csv("animals.csv")
df
```

```
##      animal count
## 1 elephant     3
## 2      cat     2
## 3     frog     6
```

```
str(df)
```

```
## 'data.frame':   3 obs. of  2 variables:
## $ animal: chr  "elephant" "cat" "frog"
## $ count : int  3 2 6
```

### 3. Base R

Base R used to prefer factors

```
df <- read.csv("animals.csv")
df
```

```
##      animal count
## 1 elephant     3
## 2      cat     2
## 3     frog     6
```

```
str(df)
```

```
## 'data.frame':    3 obs. of  2 variables:
## $ animal: chr  "elephant" "cat" "frog"
## $ count : int  3 2 6
```

```
df <- read.csv("animals.csv", stringsAsFactors = TRUE)
df
```

```
##      animal count
## 1 elephant     3
## 2      cat     2
## 3     frog     6
```

```
str(df)
```

```
## 'data.frame':    3 obs. of  2 variables:
## $ animal: Factor w/ 3 levels "cat","elephant",...: 2 1 3
## $ count : int  3 2 6
```

### 3. Tidyverse ... avoids factors

Tidyverse:

```
library(readr)
df <- read_csv("animals.csv")
str(df)
```

```
## spec_tbl_df [3 × 2] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ animal: chr [1:3] "elephant" "cat" "frog"
## $ count : num [1:3] 3 2 6
## - attr(*, "spec")=
## .. cols(
## ..   animal = col_character(),
## ..   count = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

## 4. Base R

Base:

```
df <- read.csv("animals.csv")  
df
```

```
##      animal count  
## 1 elephant      3  
## 2      cat      2  
## 3      frog      6
```

```
x <- cbind(df[,1], df[,2])  
class(x)
```

## 4. Base R

Base:

```
df <- read.csv("animals.csv")  
df
```

```
##      animal count  
## 1 elephant     3  
## 2      cat     2  
## 3     frog     6
```

```
x <- cbind(df[,1], df[,2])  
class(x)
```

```
## [1] "matrix" "array"
```

```
x
```

## 4. Base R

Base:

```
df <- read.csv("animals.csv")  
df
```

```
##      animal count  
## 1 elephant     3  
## 2      cat     2  
## 3      frog     6
```

```
df[,1]
```

```
## [1] "elephant" "cat"      "frog"
```

```
class(df[,1])
```

```
## [1] "character"
```



## 4. Tidyverse ... avoids dropping dimensions

Tidyverse:

```
tib <- read_csv("animals.csv")
tib
```

```
## # A tibble: 3 × 2
##   animal    count
##   <chr>    <dbl>
## 1 elephant     3
## 2 cat          2
## 3 frog         6
```

```
tib[,1]
```

```
## # A tibble: 3 × 1
##   animal
##   <chr>
## 1 elephant
## 2 cat
## 3 frog
```

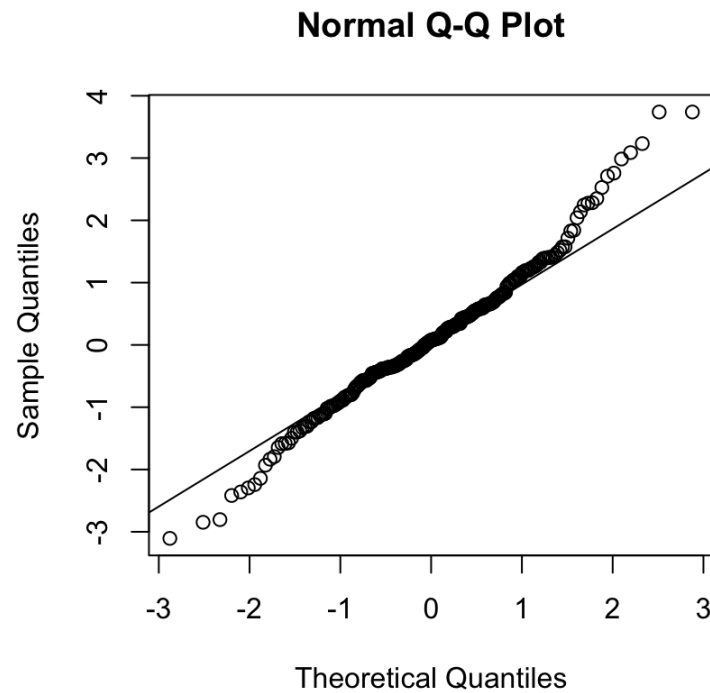
```
class(tib[,1])
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

## 5. Base R

Base:

```
# p. 115, Modern Applied Statistics with S-Plus (1999)
x <- rt(250, 9)
qqnorm(x); qqline(x)
```



Source: Venables and Ripley, *Modern Applied Statistics with S-Plus* (1999), p. 115.

(btw 1999 is “new”...)



## 5. Tidyverse is ...

Tidyverse:

```
df <- iris %>% dplyr::add_rownames()
```

## 5. Tidyverse is ... still evolving

Tidyverse:

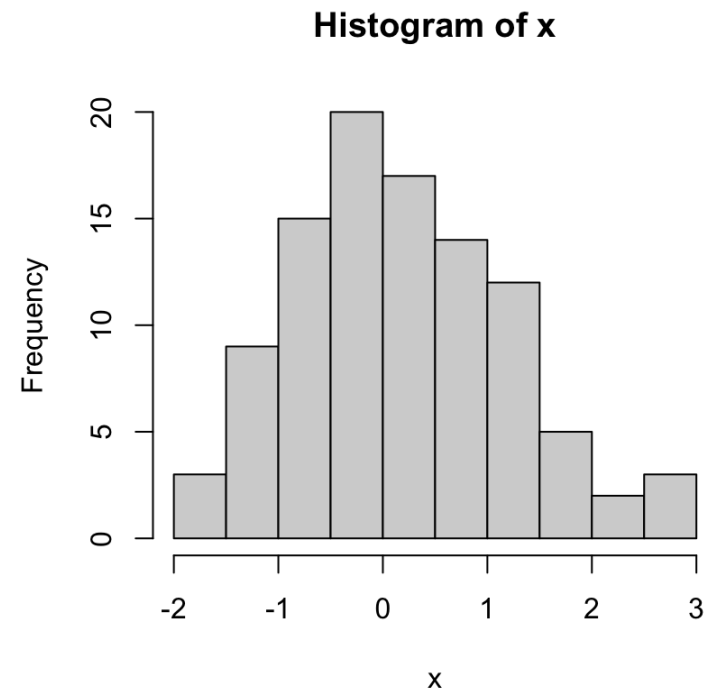
```
df <- iris %>% dplyr::add_rownames()
```

```
## Warning: `add_rownames()` was deprecated in dplyr 1.0.0.  
## Please use `tibble::rownames_to_column()` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

## 6. Base R

Base:

```
x <- rnorm(100)  
hist(x)
```



## 6. Tidyverse ... avoids vectors

Tidyverse:

```
library(ggplot2)
ggplot(x, aes(x)) + geom_histogram()
```

```
## Error: `data` must be a data frame, or other object coercible by `fortify()`, not a numeric vector.
```

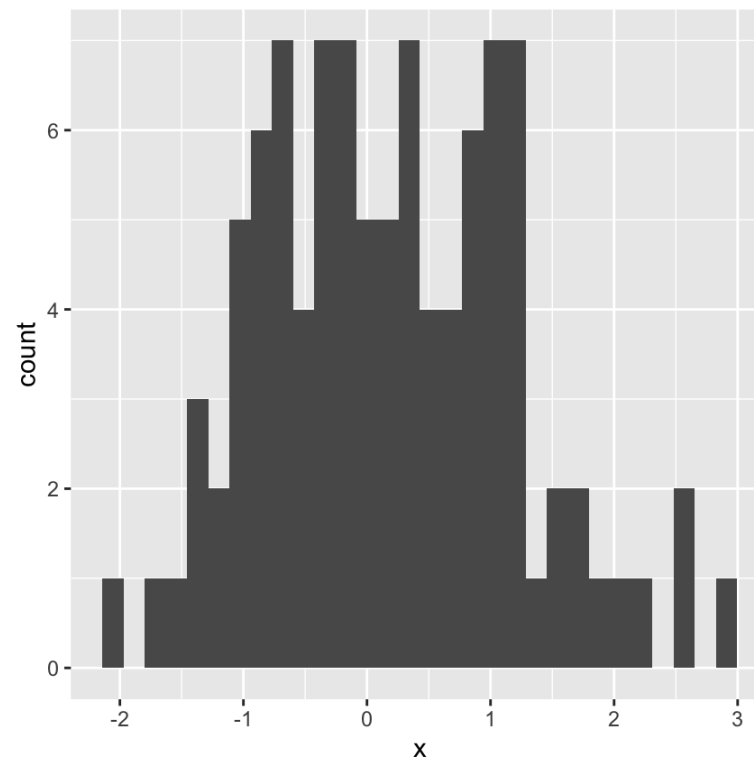
## 6. Tidyverse ... avoids vectors

Tidyverse:

```
library(ggplot2)
ggplot(x, aes(x)) + geom_histogram()
```

```
## Error: `data` must be a data frame, or other object coercible by `fortify()`, not a numeric vector.
```

```
ggplot(data.frame(x), aes(x)) + geom_histogram()
```

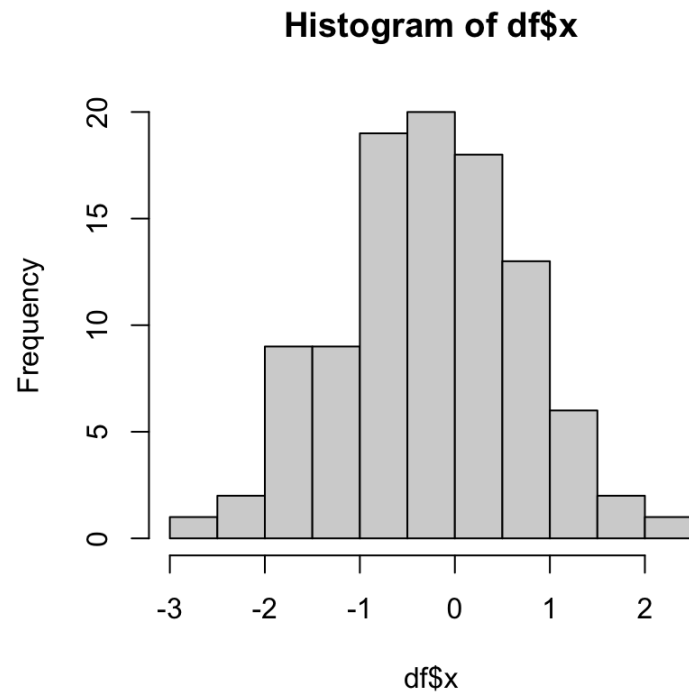




## 7. Base R

Base:

```
df <- data.frame(x = rnorm(100))  
hist(df$x)
```

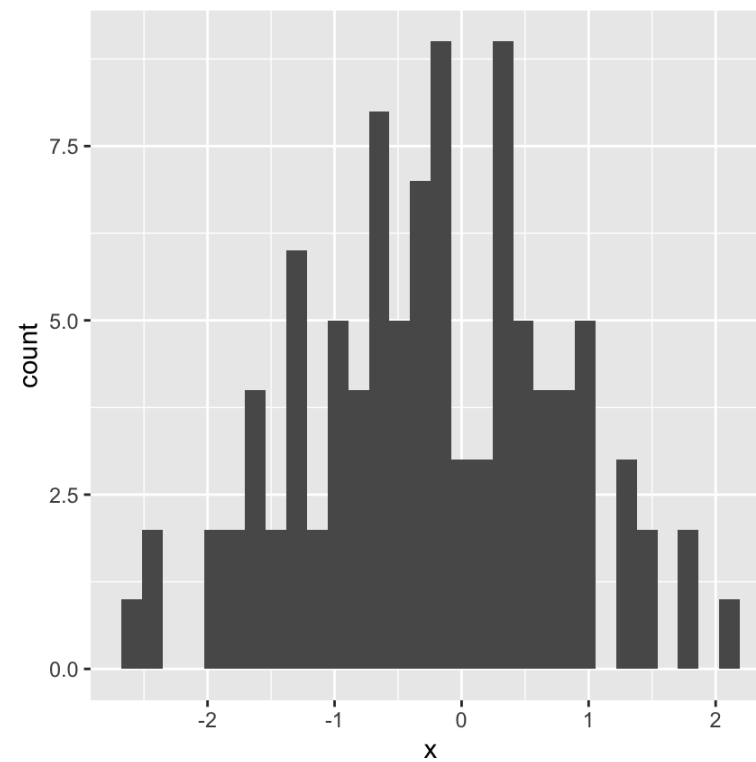


## 7. Tidyverse is ... more talkative

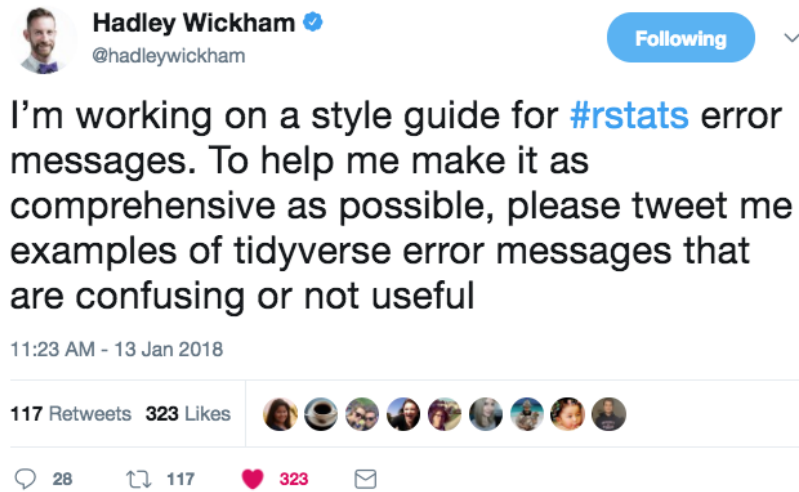
Tidyverse:

```
library(ggplot2)
ggplot(df, aes(x)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

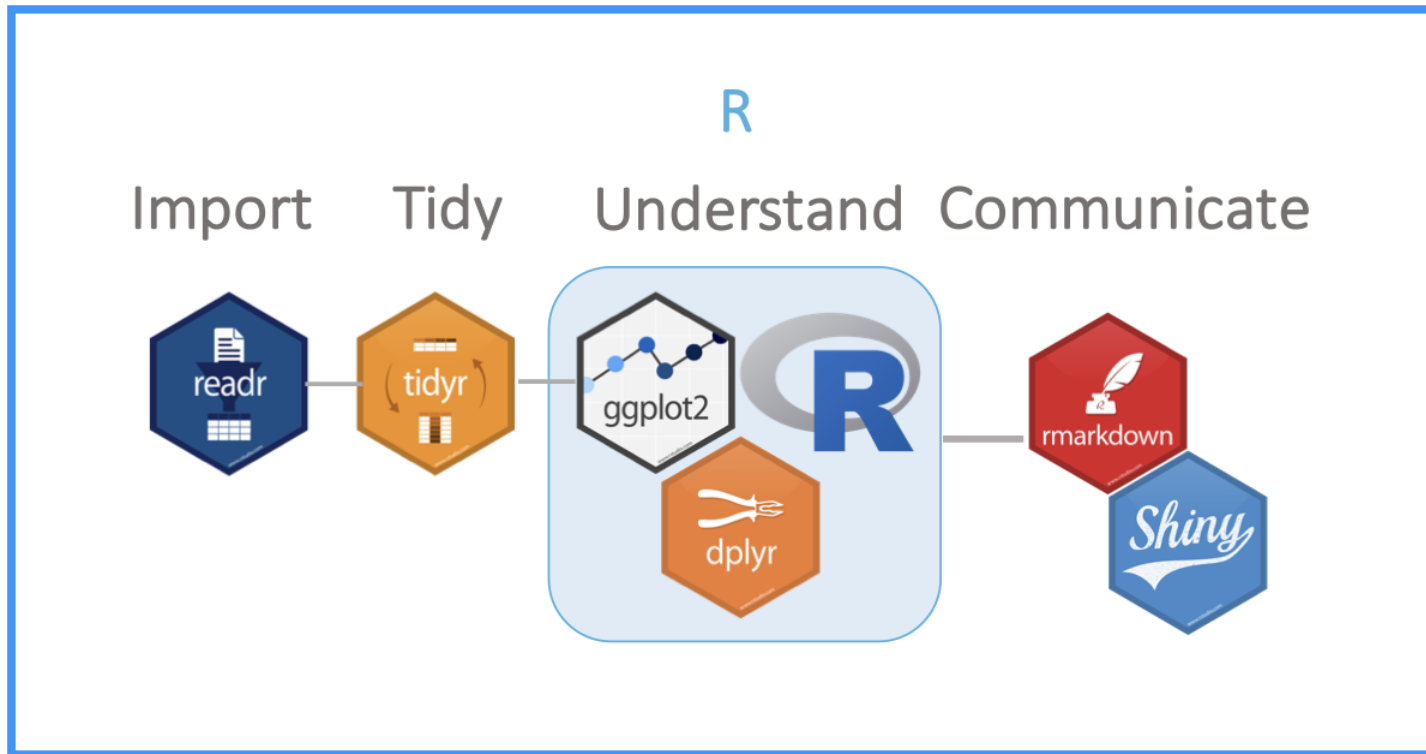


## 7. Tidyverse is ... more talkative



<https://twitter.com/hadleywickham/status/952259891342794752>

## 8. Tidyverse is .. more coordinated across tasks



Source: RStudio,

[https://github.com/rstudio/meetup\\_roadshow/blob/master/2017%20Meetup%20Roadshow.pptx](https://github.com/rstudio/meetup_roadshow/blob/master/2017%20Meetup%20Roadshow.pptx)

## 9. Tidyverse is ... easier for beginners

```
# base R
crime.by.state <- read.csv("CrimeStatebyState.csv")
crime.ny.2005 <- crime.by.state[crime.by.state$Year==2005 &
                                crime.by.state$State=="New York",
                                c("Type.of.Crime", "Count")]
crime.ny.2005 <- crime.ny.2005[order(crime.ny.2005$Count,
                                     decreasing=TRUE), ]
crime.ny.2005$Proportion <- crime.ny.2005$Count /
                             sum(crime.ny.2005$Count)
summary1 <- aggregate(Count ~ Type.of.Crime,
                      data=crime.ny.2005, FUN=sum)
summary2 <- aggregate(Count ~ Type.of.Crime,
                      data=crime.ny.2005, FUN=length)
final <- merge(summary1, summary2, by="Type.of.Crime")
```

```
# dplyr
crime.by.state <- read.csv("CrimeStatebyState.csv")
final <- crime.by.state %>%
  filter(State=="New York", Year==2005) %>%
  arrange(desc(Count)) %>%
  select(Type.of.Crime, Count) %>%
  mutate(Proportion=Count/sum(Count)) %>%
  group_by(Type.of.Crime) %>%
  summarise(num.types = n(), counts = sum(Count))
```

Goodbye \$ [ ]

Source: “How dplyr replaced my most common R idioms”

<http://www.onthelambda.com/2014/02/10/how-dplyr-replaced-my-most-common-r-idioms/>

(highly recommended!)

## 10. Tidyverse ... is more collaborative

The screenshot shows the GitHub repository page for `tidyverse/forcats`. The repository is public and has 22 watchers, 124 forks, and 464 stars. The main branch is `main`, with 3 branches and 7 tags. The repository description is: "🐼🐼🐼: tools for working with categorical variables (factors)". The repository includes a README, a license, and a code of conduct. The repository has 464 stars, 22 watching, and 124 forks. The repository also has 7 releases, with the latest release being `forcats 0.5.1`.

Repository: `tidyverse/forcats` (Public)

Watch 22 Fork 124 Starred 464

Code Issues 35 Pull requests 3 Actions Security Insights

main 3 branches 7 tags

Go to file Add file Code

**Commits**

Commit	Message	Time
jennybc	Update cheatsheet links	303 commits
b4dade0	on Nov 29, 2021	
.github	Update GHA	3 months ago
R	Fix doc (#278)	14 months ago
data-raw	Add some continuous vars too	6 years ago
data	Add some continuous vars too	6 years ago
man	Update for master --> main	3 months ago
pkgdown/favicon	Prepare for autodeploy	3 years ago
revdep	Revert #237 to avoid haven issue	2 years ago
tests	Revert #237 to avoid haven issue	2 years ago
vignettes	Update urls	12 months ago
...Rbuildignore	Continuous integration with GitHub Actions (#246)	2 years ago

**About**

🐼🐼🐼: tools for working with categorical variables (factors)

[forcats.tidyverse.org](https://forcats.tidyverse.org)

r tidyverse factor

Readme View license Code of conduct

464 stars 22 watching 124 forks

**Releases** 7

forcats 0.5.1 (Latest)