

Cleaning Data

Kathleen Zhang

Read in the data

```
library(tidyverse)
library(tidymodels)
library(dplyr)
library(lubridate)

getwd()

[1] "/Users/kathleenzhang/Diverse Led Funds/10_code"

setwd("/Users/kathleenzhang/Diverse Led Funds/00_source_data")

basediversity <- read.csv("baseventures.csv")
base <- read.csv("baseventures2.csv")
```

Cleaning Data

```
basediversity["Money.Raised"][basediversity["Money.Raised"] == "-"] <- NA

as.numeric(basediversity$Announced.Date)

[1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[26] NA NA NA NA NA NA NA NA NA NA

numericaldate <- lubridate::dmy(basediversity$Announced.Date)

basediversity |>
  mutate(year = as.Date(basediversity$Announced.Date, "%m/%d/%Y"), "%Y")
```

	Announced.Date	Organization.Name
1	Oct 14, 2021	Space Perspective
2	Nov 9, 2020	Space Perspective
3	Apr 17, 2019	Mos
4	Mar 1, 2018	Frank Darling
5	Sep 19, 2017	May Mobility
6	Apr 3, 2017	WearAway
7	Sep 20, 2016	O.School
8	Sep 13, 2016	Blavity
9	Jul 5, 2016	Netki
10	May 16, 2016	FleetPanda
11	May 1, 2016	TravelJoy
12	Apr 28, 2016	World View Enterprises
13	Apr 7, 2016	Luma
14	Mar 7, 2016	SuperPhone
15	Aug 5, 2015	Flip
16	Nov 17, 2014	OrderWithMe
17	Nov 6, 2014	Naja
18	Oct 10, 2014	Monsieur
19	Jun 2, 2014	Mayvenn
20	May 20, 2014	6sense
21	May 14, 2014	Cognea
22	Feb 17, 2014	PandoDaily
23	Jan 22, 2014	StyleSeat
24	Jan 8, 2014	Shelf
25	Dec 19, 2013	Thinair
26	Dec 4, 2013	OrderWithMe
27	Nov 27, 2013	Wildfang
28	Oct 15, 2013	World View Enterprises
29	Sep 20, 2013	Pigeonly
30	Aug 1, 2012	ClassBadges
31	May 29, 2012	PlanGrid
32	Apr 30, 2012	Socialcam
33	Apr 1, 2011	Relevant
34	Sep 10, 2010	inDinero

Diversity.Spotlight..US.Hea

1		Women Fou
2		Women Fou
3	Middle Eastern / North African Founded, Middle Eastern / North African Led,	Women Fou
4		Women Fou
5		
6		Women Fou
7		Women Fou

8 Black / African American Founded, Black / African American Led, Women Fou
9
10 South
11 Black / African American Founded, Black / Afric
12
13 Black / African A
14 Black / African American Founded, Black / Afric
15 Black / African American Founded, Black / Afric
16
17 Hispanic / Latine Founded, Hispanic / Latine Led, Women Fou
18 Black / African American Founded, Black / Afric
19 Black / African American Founded, Black / Afric
20
21 Women Fou
22 Women Fou
23 Women Fou
24 Black / African American Founded
25 Black / African A
26
27 Women Fou
28
29 Black / African American Founded, Black / Afric
30
31
32 Black / African A
33 Black / African A
34

	Funding.Round	Money.Raised	year	"%Y"
1	Series A - Space Perspective	\$41M	<NA>	%Y
2	Seed Round - Space Perspective	\$7M	<NA>	%Y
3	Seed Round - Mos	\$4M	<NA>	%Y
4	Pre Seed Round - Frank Darling	<NA>	<NA>	%Y
5	Seed Round - May Mobility	<NA>	<NA>	%Y
6	Seed Round - WearAway	<NA>	<NA>	%Y
7	Pre Seed Round - O.School	<NA>	<NA>	%Y
8	Seed Round - Blavity	<NA>	<NA>	%Y
9	Seed Round - Netki	\$3.5M	<NA>	%Y
10	Seed Round - FleetPanda	\$125K	<NA>	%Y
11	Pre Seed Round - TravelJoy	<NA>	<NA>	%Y
12	Series B - World View Enterprises	\$15M	<NA>	%Y
13	Series A - Luma	\$12.5M	<NA>	%Y
14	Seed Round - SuperPhone	\$1.5M	<NA>	%Y
15	Seed Round - Flip	<NA>	<NA>	%Y

16	Series B - OrderWithMe	\$28M <NA>	%Y
17	Seed Round - Naja	<NA> <NA>	%Y
18	Seed Round - Monsieur	\$2M <NA>	%Y
19	Seed Round - Mayvenn	<NA> <NA>	%Y
20	Series A - 6sense	\$12M <NA>	%Y
21	Seed Round - Cognea	<NA> <NA>	%Y
22	Series A - PandoDaily	\$1.2M <NA>	%Y
23	Series A - StyleSeat	\$10.2M <NA>	%Y
24	Seed Round - Shelf	\$1.2M <NA>	%Y
25	Seed Round - Thinair	<NA> <NA>	%Y
26	Series B - OrderWithMe	\$6M <NA>	%Y
27	Venture Round - Wildfang	\$2M <NA>	%Y
28	Seed Round - World View Enterprises	\$250K <NA>	%Y
29	Seed Round - Pigeonly	\$1M <NA>	%Y
30	Seed Round - ClassBadges	<NA> <NA>	%Y
31	Seed Round - PlanGrid	\$1.1M <NA>	%Y
32	Angel Round - Socialcam	<NA> <NA>	%Y
33	Convertible Note - Relevvant	\$675K <NA>	%Y
34	Seed Round - inDinero	\$1.2M <NA>	%Y

```

keeps <- c("PORTFOLIO.COMPANY", "STAGE", "DEAL.DATE", "DEAL.SIZE..CURR..MN.", "PORTFOLIO.SECTOR")

base = base[keeps]

```

Graphing Data

- `ggplot(data = basediversity, mapping = aes(x = "Diversity.Spotlight", y = "Money.Raised", color = "Sector")) + geom_col()`

