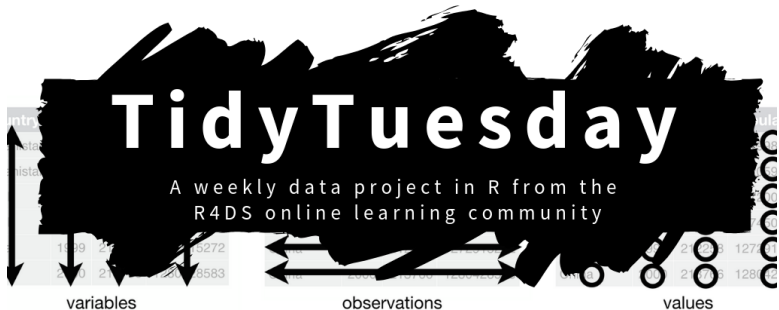


Fun with Bob Ross and #TidyTuesday
(HARUG! 2020.03.04)

Ed Harris

02/03/2020

Tidy Tuesday



#TidyTuesday is an activity to build your **data handling** and **R graphics** skills. Each week (on Tuesday) a new dataset is posted. The task is to tidy up the data and make a graph that tells a story. The “#” is a Twitter hashtag. People use this to Tweet their creative graphs.

#TidyTuesday links

Official site <https://github.com/rfordatascience/tidytuesday>

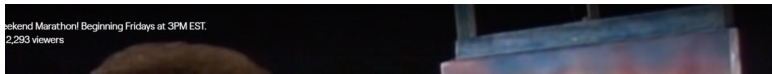
Datasets and info (thanks Joe Roberts)

<https://nsgrantham.shinyapps.io/tidytuesdayrocks/>

Bob Ross



Weekend Marathon! Beginning Fridays at 3PM EST.
2,293 viewers



Live coding

The screenshot displays a live coding session in RStudio. The left pane shows R code for performing PCA on painting elements. The right pane shows four bar charts representing the first four principal components (PC1-PC4) for the elements in Bob Ross paintings.

```
150 coord_flip() +
151 labs(title = "First four principal components of elements in Bob Ross paintings")
152
153
154 1. Mountains/Conifer vs Ocean/Beach and deciduous trees
155 2. Trees, especially deciduous, vs Ocean
156 3. Spring/Summer vs Winter
157 4. Lake vs River
158
159 "" [r]
160 painting_weights <- broom::tidy(svd_result, matrix = "w") %>%
161   mutate(painting = rownames(binary_matrix)[row])
162
163 "" [r]
164
165 painting_weights %>%
166   filter(PC == 1) %>%
167   arrange(desc(value))
168
169 bob_ross_gathered %>%
170   gather(element, value, PC1:PC4)
```

The console shows the following data table:

episode	season	episode_number	title	element
1	1	1	Mountain Waterfall	Deciduous
2	1	2	Mountain Waterfall	Grass
3	1	3	Mountain Waterfall	Lake
4	1	4	Mountain Waterfall	Mountain
5	1	5	Mountain Waterfall	Mountains
6	1	6	Mountain Waterfall	Snowy Mountain
7	1	7	Mountain Waterfall	Tree
8	1	8	Mountain Waterfall	Trees
9	1	9	Mountain Waterfall	Waterfall
10	1	10	Mountain Waterfall	Bushes
11	1	11	Mountain Waterfall	Clouds
12	1	12	Mountain Waterfall	Conifer
13	1	13	Mountain Waterfall	Cumulus
14	1	14	Mountain Waterfall	Lake
15	1	15	Mountain Waterfall	Mountain
16	1	16	Mountain Waterfall	Mountains
17	1	17	Mountain Waterfall	Snowy Mountain
18	1	18	Mountain Waterfall	Tree
19	1	19	Mountain Waterfall	Trees

The four bar charts show the following elements and their approximate values for PC1, PC2, PC3, and PC4:

- PC1:** Mountain, Conifer, Snowy Mountain, Mountains, Lake, Trees, Tree, Bushes, Winter, Snow, Mountain, Rocks, Beach, Deciduous, Waves, Ocean.
- PC2:** Deciduous, Tree, Trees, Structure, Cabin, Grass, Snow, Mountains, Snowy Mountain, Mountain, Rocks, Beach, Cumulus, Waves, Clouds.
- PC3:** River, Grass, Bushes, Deciduous, Waterfall, Rocks, Mountain, Trees, Lake, Fence, Conifer, Cabin, Structure, Conifer, Snow.
- PC4:** Lake, Deciduous, Cumulus, Clouds, Bushes, Path, Grass, Structure, Cabin, Snowy Mountain, Waterfall, Snow, Winter, Conifer, River.

<https://www.youtube.com/watch?v=sD993H5FBIY>

Now we try

- 1) download Bob Ross data
- 2) explore the data
- 3) formulate 1 (of many possible) question
- 4) make a first graph
- 5) feedback, critique and make it better > new graph
- 6) repeat 5 until "done"

Task for next time

- 1) get any #TidyTuesday dataset
- 2) build a first graph (TM) and share it in slack
https://join.slack.com/t/harper-adams-rug/shared_invite/enQtODYwNTkyMDMzODkyLTVhYzYzZDI1N2JjMGZlZyBkNmQ0M
(get feedback, ask advice)
- 3) make the graph better
- 4) repeat #3 until you are proud of it :)