

Wine Reviews

Joel Soroos

May 31, 2019

1. Get data

```
library (tidyverse)
library (janitor)

wine_raw <- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/wine/wine.csv")
clean_names ()
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

2. Transform data

```
wine_raw %>%
  group_by (variety) %>%
  summarize (n = n()) %>%
  arrange (-n) %>%
  top_n (10) %>%
  select (-n) -> wine_group

wine_raw %>%
  drop_na (variety, price) %>%
  mutate (
    variety = as.factor(variety),
    popularity = ifelse (variety %in% wine_group$variety, "High", "Low"),
    color = case_when(
      variety %in% c("Pinot Noir", "Cabernet Sauvignon", "Red Blend", "Bordeaux-style Red Blend",
        variety %in% c("Pinot Noir", "Chardonnay", "Riesling", "Sauvignon Blanc") ~ "Oldlace",
        variety %in% c("Rosé") ~ "Deeppink"
      )
    ) %>%
  filter (
    popularity == "High",
    price < 300
  ) %>%
  select (variety, color, price) -> wine
```

3. Visualize data

```
library (ggdark)
library (scales)
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##   discard

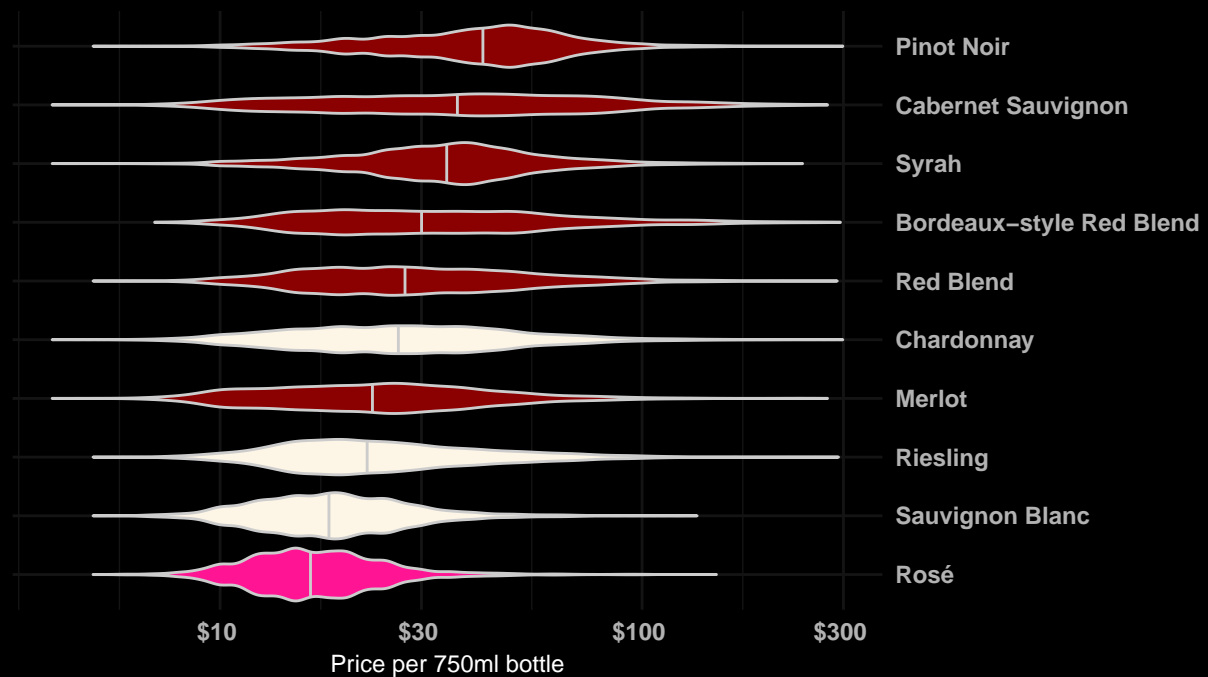
## The following object is masked from 'package:readr':
##
##   col_factor
```

```
ggplot (
  data = wine,
  aes (x = reorder(variety, price, median), y = price, fill = I(color))
) +
geom_violin(draw_quantiles = c(.50)) +
scale_x_discrete(position = "top") +
scale_y_continuous (
  trans = log10_trans(),
  label = unit_format(prefix = "$", unit = "")
) +
coord_flip () +
dark_mode(theme_minimal()) +
theme (
  plot.title = element_text(hjust = 0, size = 15),
  plot.caption = element_text(hjust = 0, size = 9),
  axis.title=element_text(size=9),
  axis.text=element_text(size=9, face = "bold"),
  axis.ticks = element_blank()
) +
labs(
  title = "Open your wallet wide if you prefer red wine!",
  caption = "\nVertical line within violin represents median wine bottle price for that grape va",
  x = "",
  y = "Price per 750ml bottle"
) +
ggsave("wine.png")
```

```
## Inverted geom defaults of fill and color/colour.
## To change them back, use invert_geom_defaults().
```

```
## Saving 6.5 x 4.5 in image
```

Open your wallet wide if you prefer red wine!



Vertical line within violin represents median wine bottle price for that grape variety.

Varietals with 10 most reviews displayed. Excludes wines priced > \$300 per bottle.

Source: Wine Enthusiast via Kaggle via R4DS Tidy Tuesday | Visualization: Joel Soroos @soroosj