

Museums of New York City

Data Scraping

```
library(tidyverse)
```

```
-- Attaching packages ----- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   0.3.5
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(rvest)
```

Attaching package: 'rvest'

The following object is masked from 'package:readr':

guess_encoding

```
library(lubridate)
```

Loading required package: timechange

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

date, intersect, setdiff, union

```
library(robotstxt)
library(scales)
```

Attaching package: 'scales'

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

discard

The following object is masked from 'package:readr':

col_factor

```
paths_allowed("https://museumhack.com/museums-in-nyc/#list-museums-in-NYC")
```

museumhack.com

[1] TRUE

```
page <- read_html("https://museumhack.com/museums-in-nyc/#list-museums-in-NYC")
```

```
name <- page |>
  html_elements(".h2") |>
  html_text2()
```

```
category <- page |>
  html_elements(".text-muted") |>
  html_text2() |>
  str_remove(" Museum")
```

```
price <- page |>
```

```

html_elements("li:nth-child(2) .small") |>
html_text2() |>
str_remove(" per person")

address <- page |>
  html_elements("li~ li+ li .small") |>
  html_text2()

highlight1 <- page |>
  html_elements(".col-sm-8 li:nth-child(1)") |>
  html_text2()

museums_raw <- tibble(
  name = name,
  category = category,
  price = price,
  address = address,
  highlight1 = highlight1
)

museums <- museums_raw |>
  mutate(price = str_remove(price, "\\$"),
         price = str_remove(price, " "),
         price = na_if(price, ""),
         price = na_if(price, "Donation"),
         price = case_when(price == "20-29" ~ "20",
                           price == "15+" ~ "15",
                           TRUE ~ price),
         price = as.integer(price))
glimpse(museums)

```

Rows: 145

Columns: 5

```

$ name      <chr> "The Metropolitan Museum of Art", "The American Museum of N~
$ category  <chr> "Encyclopedic", "Natural History", "Art", "Encyclopedic", "~
$ price     <int> 25, 23, 19, 16, 25, 7, 26, 22, 18, 25, 25, 22, 20, 20, 33, ~
$ address   <chr> "1000 5th Ave, New York, NY 10028", "Central Park West & 79~
$ highlight1 <chr> "Temple of Dendur", "Lucy", "The Wish Fulfilling Tree", "Co~

```

```

meanprice <- museums |>
  group_by(category) |>
  na.omit() |>
  mutate(catmeanprice = mean(price),
         catmeanprice = as.integer(catmeanprice))

meanprice <- meanprice |>
  group_by(category) |>
  mutate(pricedifference = price - catmeanprice)

catmeanprice <- museums |>
  drop_na() |>
  group_by(category) |>
  summarise(meanprice = mean(price))

archaeological <- meanprice |>
  filter(category == "Archaeological")

art <- meanprice |>
  filter(category == "Art")

childrens <- meanprice |>
  filter(category == "Children's")

cultural <- meanprice |>
  filter(category == "Cultural Center")

encyclopedic <- meanprice |>
  filter(category == "Encyclopedic")

historic <- meanprice |>
  filter(category == "Historic Site")

history <- meanprice |>
  filter(category == "History")

library <- meanprice |>
  filter(category == "Library")

natural <- meanprice |>
  filter(category == "Natural History")

```

```
religious <- meanprice |>
  filter(category == "Religious")

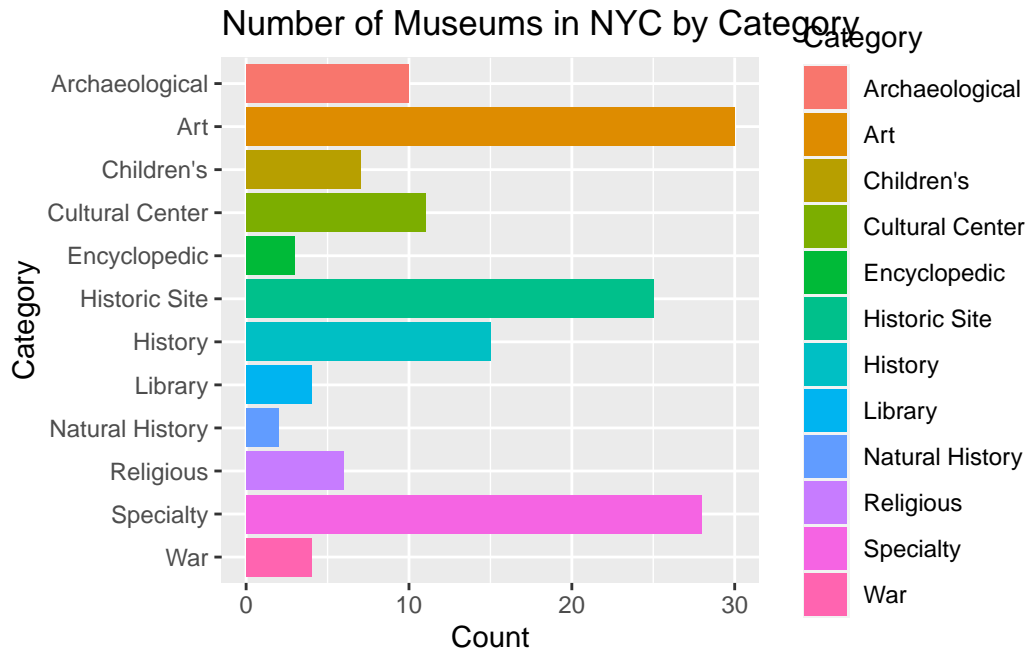
specialty <- meanprice |>
  filter(category == "Specialty")

war <- meanprice |>
  filter(category == "War")
```

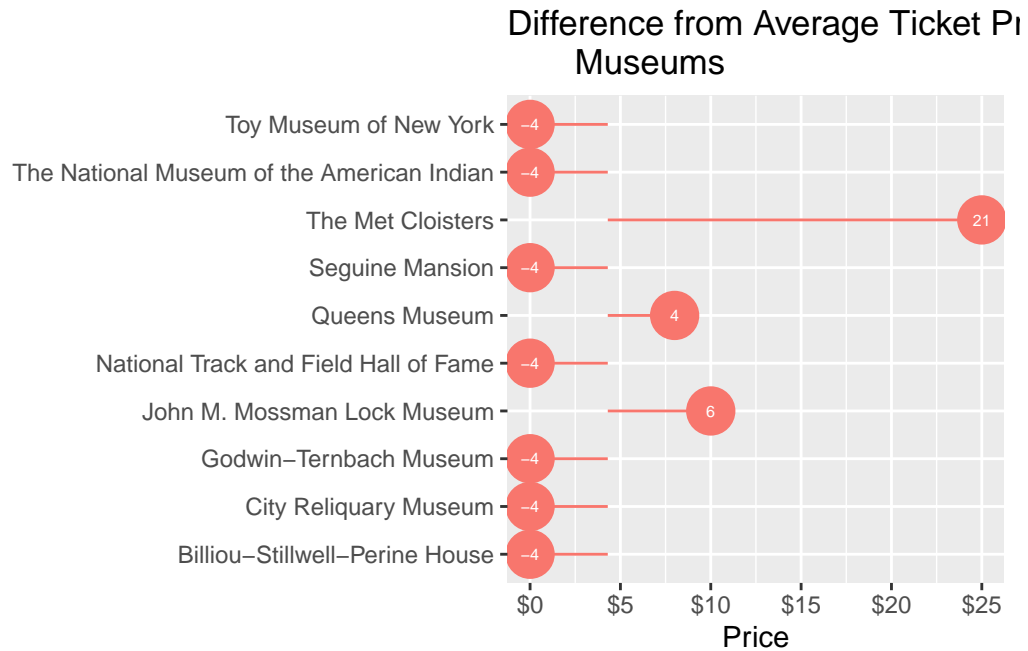
Visualizing Data

```
museumscat <- museums |>
  group_by(category) |>
  count(category)

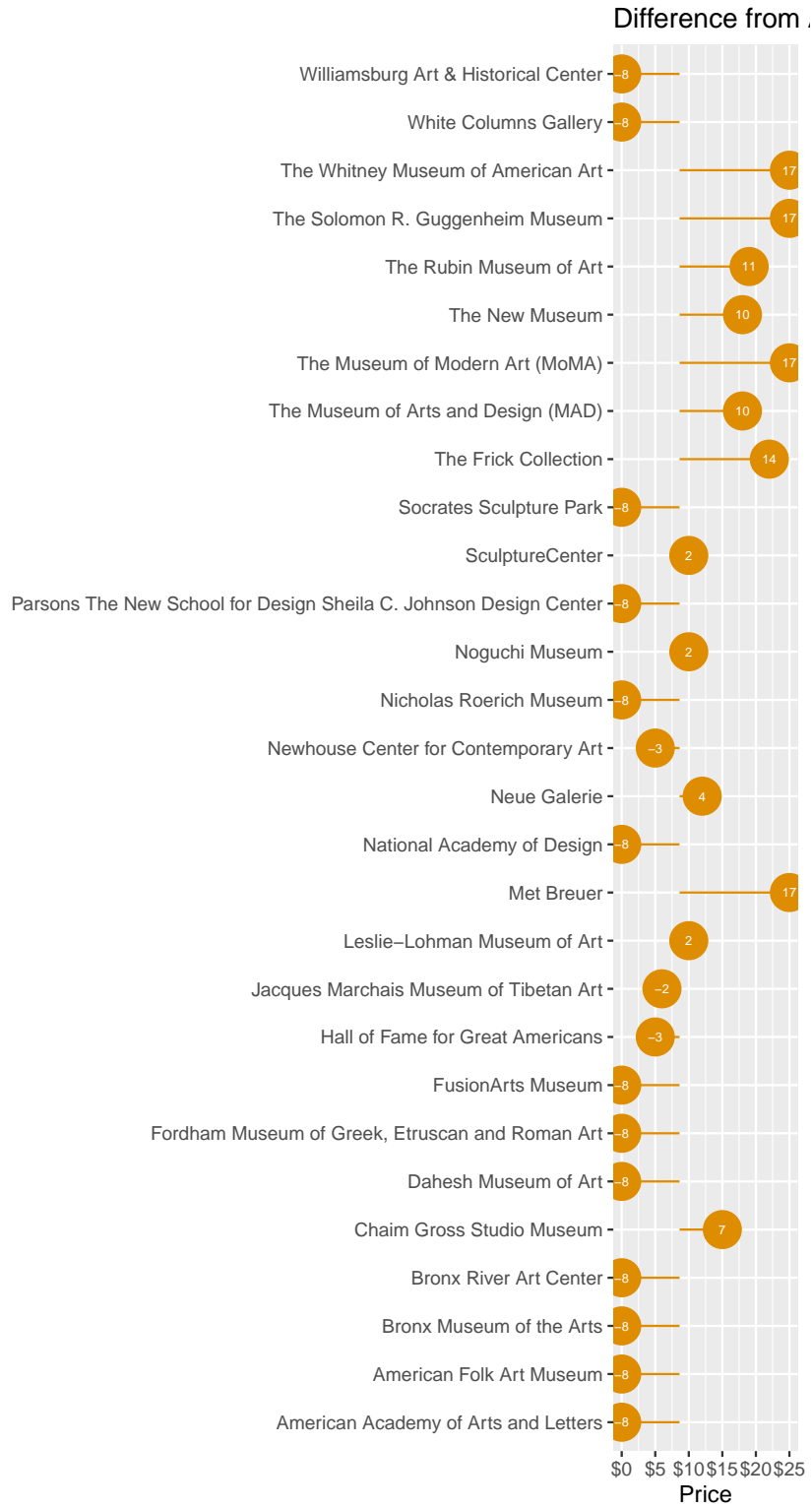
museumscat |>
  ggplot(
    aes(x = n, y = category, fill = category)) +
  geom_bar(stat = "identity") +
  labs(x = "Count",
       y = "Category",
       title = "Number of Museums in NYC by Category",
       fill = "Category") +
  scale_y_discrete(limits = rev)
```



```
archaeological |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#f8766d") +
  geom_segment(aes(y = 4.3,
    x = name,
    yend = price,
    xend = name),
    color = "#f8766d") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Archaeological
    Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")
```



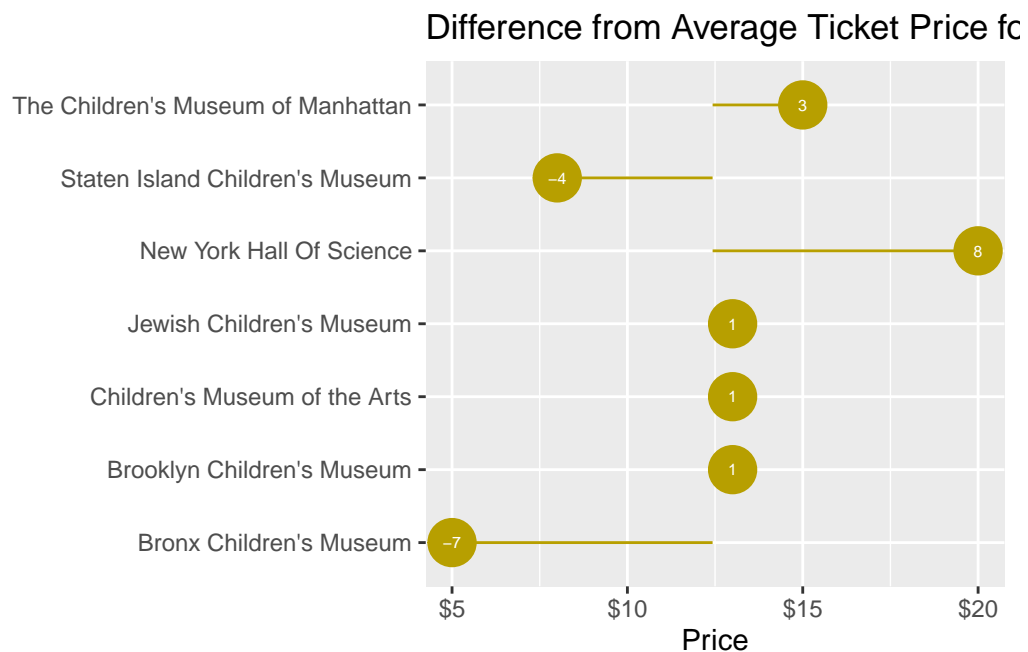
```
art |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#de8c00") +
  geom_segment(aes(y = 8.621,
    x = name,
    yend = price,
    xend = name),
    color = "#de8c00") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Art Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")
```




```

childrens |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#b79f00") +
  geom_segment(aes(y = 12.43,
    x = name,
    yend = price,
    xend = name),
    color = "#b79f00") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Children's Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")

```



```

cultural |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#7cae00") +

```

```

geom_segment(aes(y = 2.91,
                 x = name,
                 yend = price,
                 xend = name),
             color = "#7cae00") +
geom_text(color="white", size=2) +
labs(title = "Difference from Average Ticket Price for NYC Cultural Center
          Museums",
      y = "Price",
      x = "") +
scale_y_continuous(labels = scales::dollar) +
coord_flip() +
theme(legend.position="none")

```

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

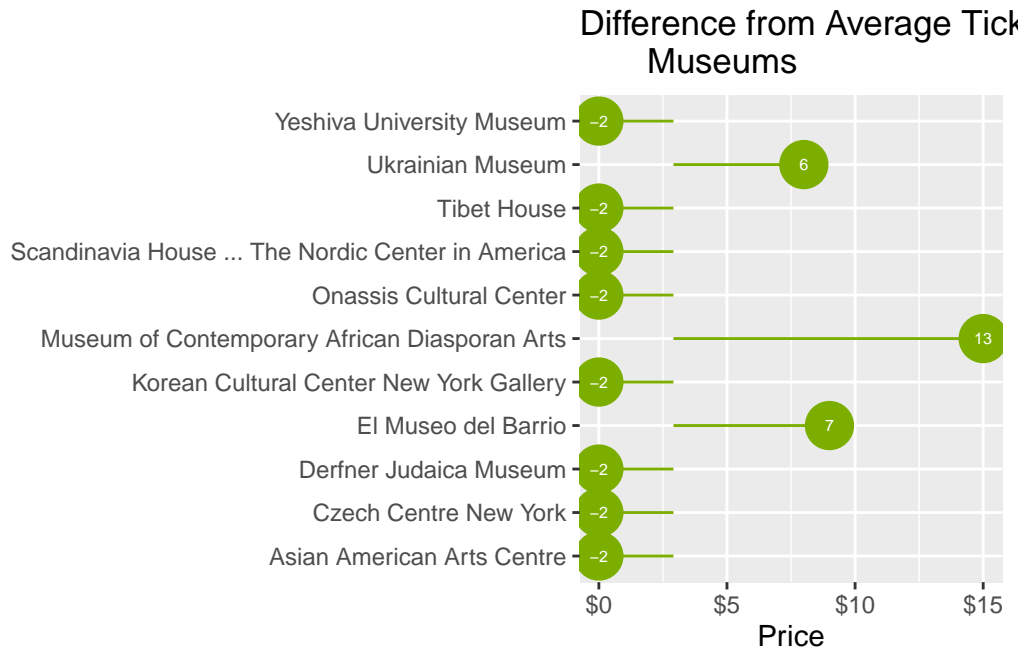
Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call(C_textBounds, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <e2>

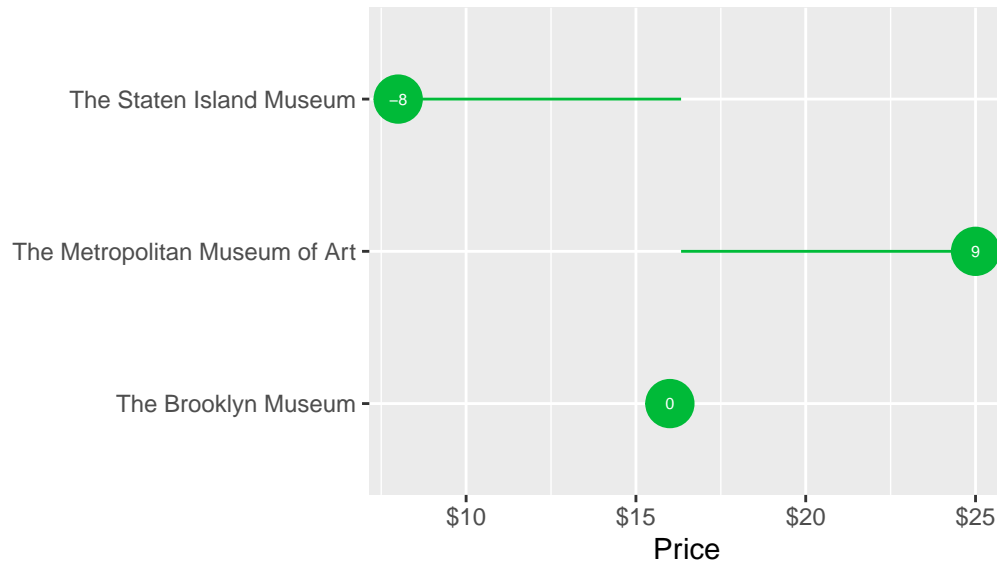
Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <80>

Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x\$label), x\$x, x\$y, :
conversion failure on 'Scandinavia House - The Nordic Center in America' in
'mbcsToSbcs': dot substituted for <93>

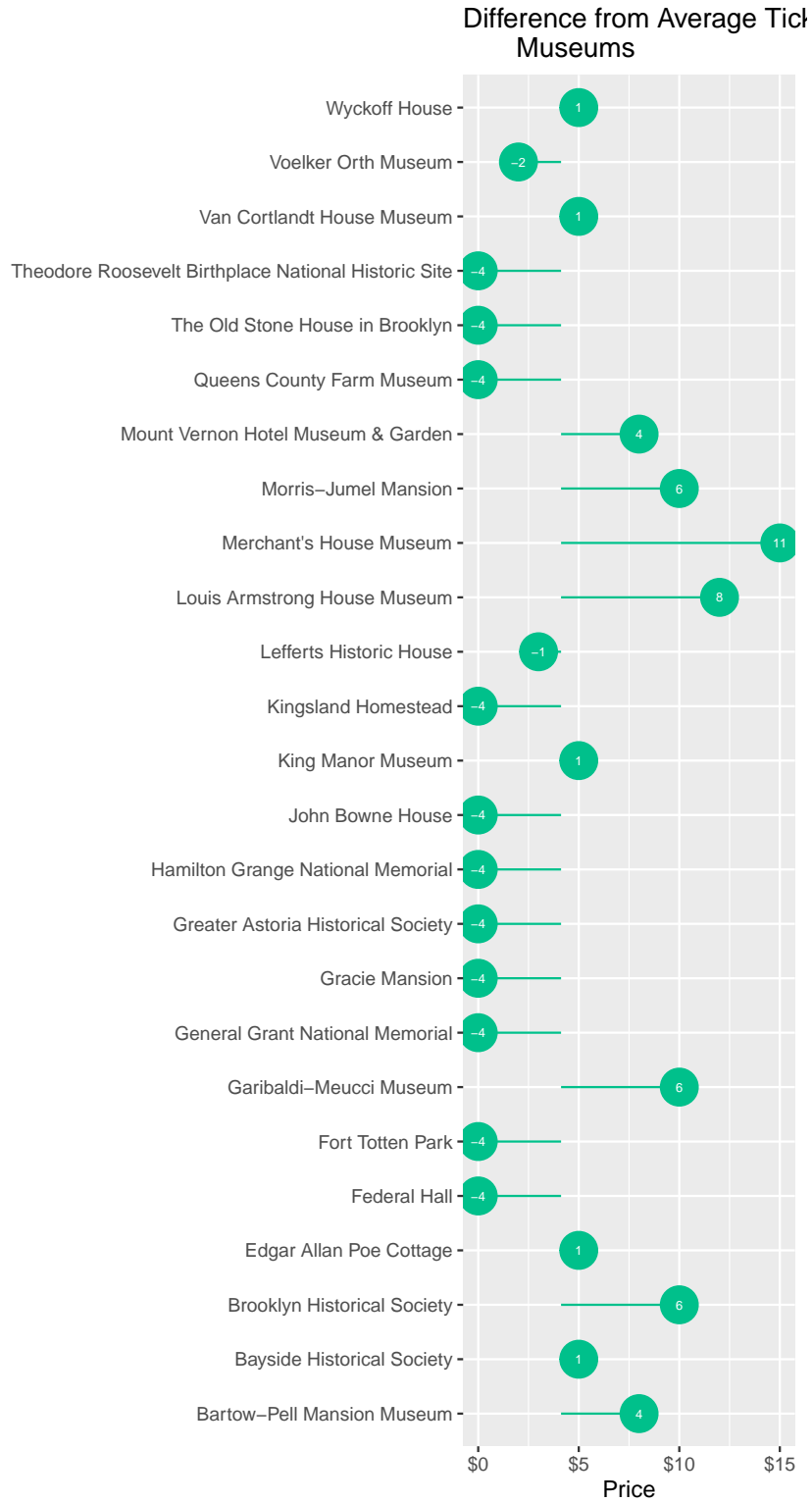


```
encyclopedic |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#00ba38") +
  geom_segment(aes(y = 16.33,
    x = name,
    yend = price,
    xend = name),
    color = "#00ba38") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Encyclopedic
    Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")
```

Difference from Average Ticket Price for N' Museums



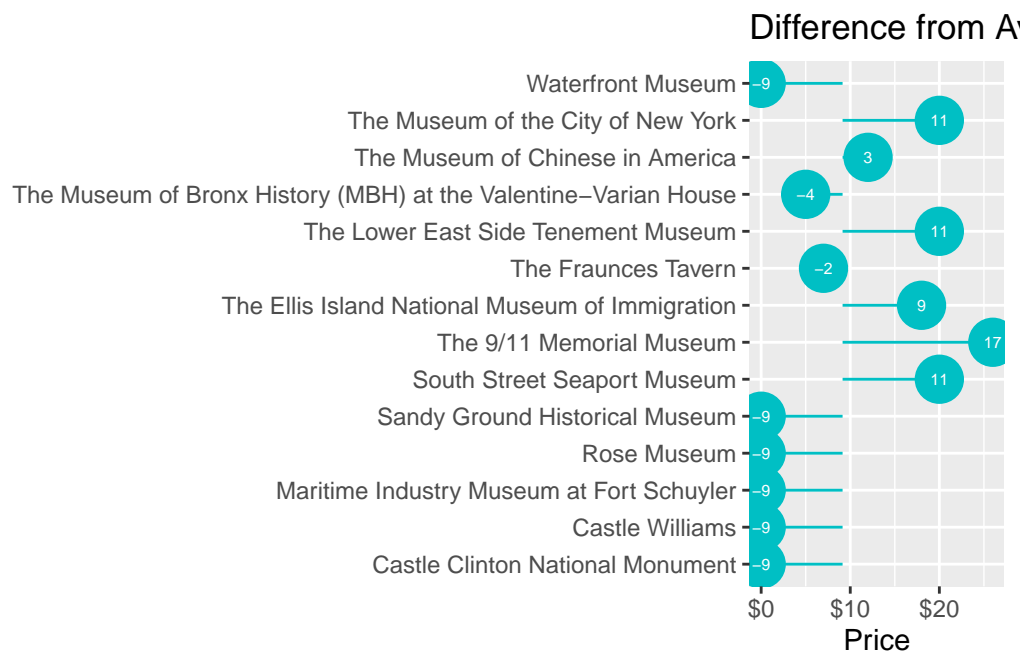
```
historic |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#00c08b") +
  geom_segment(aes(y = 4.12,
    x = name,
    yend = price,
    xend = name),
    color = "#00c08b") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Historic Site
    Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")
```

```

history |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#00bfc4") +
  geom_segment(aes(y = 9.14,
                    x = name,
                    yend = price,
                    xend = name),
               color = "#00bfc4") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC History Museums",
        y = "Price",
        x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")

```



```

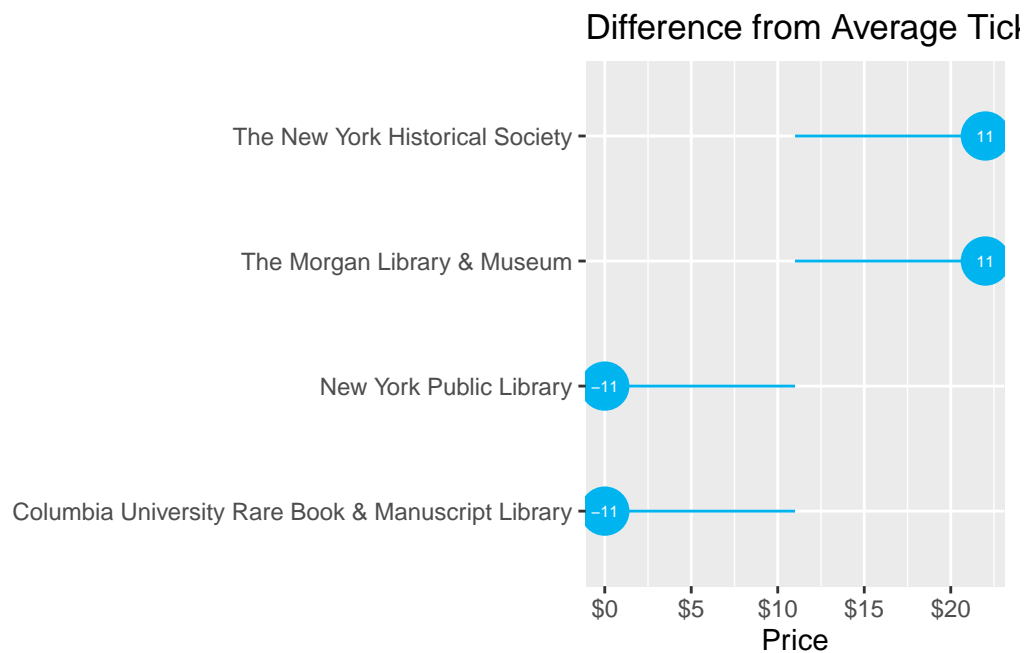
library |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#00b4f0") +

```

```

geom_segment(aes(y = 11,
                 x = name,
                 yend = price,
                 xend = name),
             color = "#00b4f0") +
geom_text(color="white", size=2) +
labs(title = "Difference from Average Ticket Price for NYC Library Museums",
     y = "Price",
     x = "") +
scale_y_continuous(labels = scales::dollar) +
coord_flip() +
theme(legend.position="none")

```



```

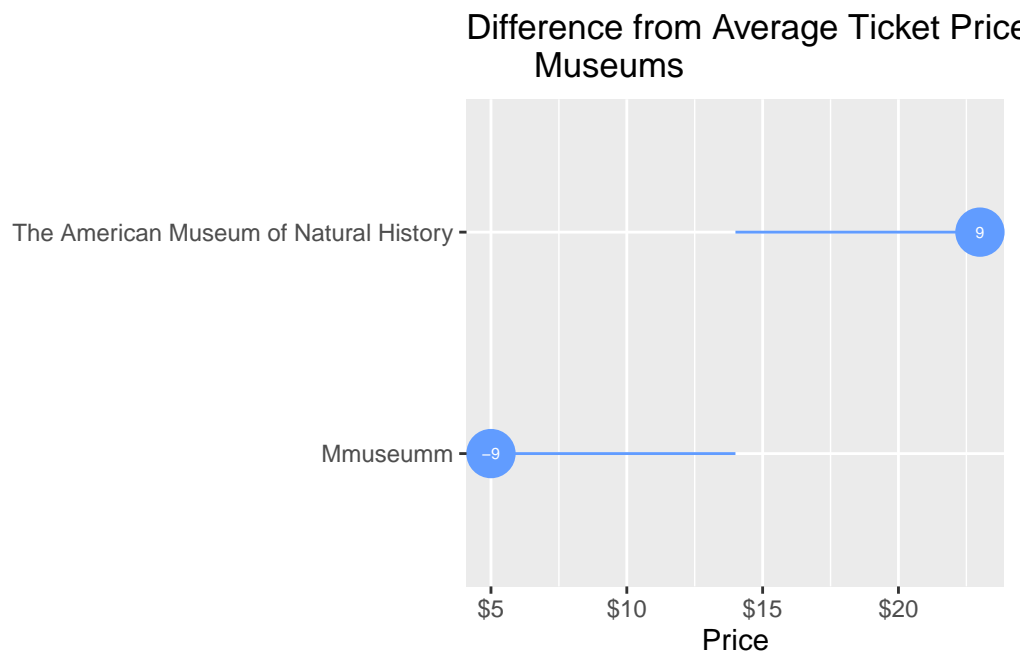
natural |>
ggplot(
  aes(x = name, y = price, label = pricedifference)) +
geom_point(size = 8, show.legend = FALSE, color = "#619cff") +
geom_segment(aes(y = 14,
                 x = name,
                 yend = price,
                 xend = name),

```

```

        color = "#619cff") +
geom_text(color="white", size=2) +
labs(title = "Difference from Average Ticket Price for NYC Natural History
Museums",
      y = "Price",
      x = "") +
scale_y_continuous(labels = scales::dollar) +
coord_flip() +
theme(legend.position="none")

```



```

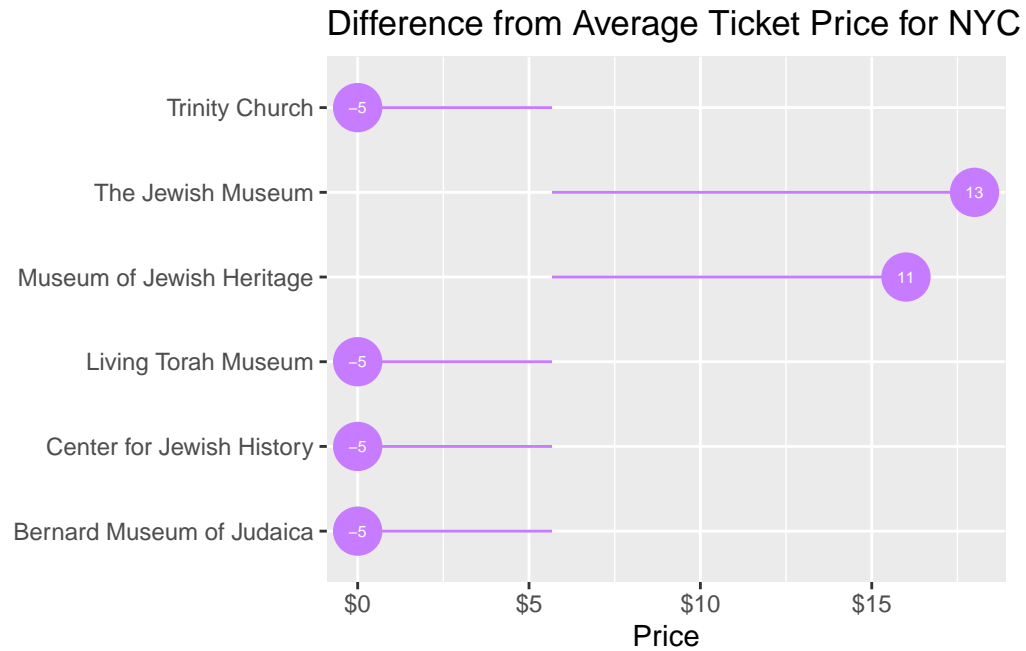
religious |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#c77cff") +
  geom_segment(aes(y = 5.67,
                    x = name,
                    yend = price,
                    xend = name),
               color = "#c77cff") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Religious Museums",

```

```

    y = "Price",
    x = "") +
scale_y_continuous(labels = scales::dollar) +
coord_flip() +
theme(legend.position="none")

```

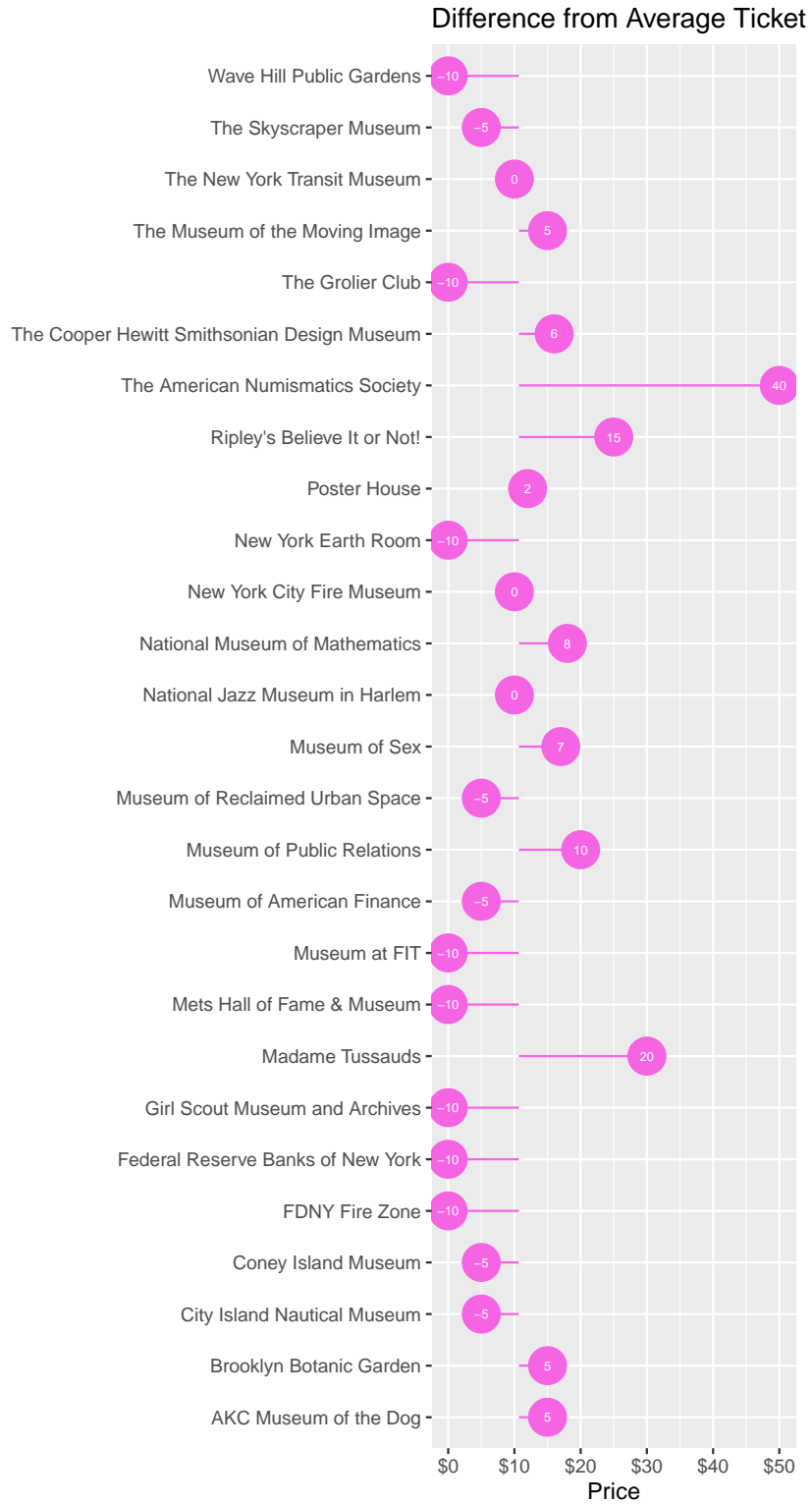


```

specialty |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#f564e3") +
  geom_segment(aes(y = 10.67,
    x = name,
    yend = price,
    xend = name),
    color = "#f564e3") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC Specialty Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +

```

```
theme(legend.position="none")
```



```

war |>
  ggplot(
    aes(x = name, y = price, label = pricedifference)) +
  geom_point(size = 8, show.legend = FALSE, color = "#ff64b0") +
  geom_segment(aes(y = 12,
    x = name,
    yend = price,
    xend = name),
    color = "#ff64b0") +
  geom_text(color="white", size=2) +
  labs(title = "Difference from Average Ticket Price for NYC War Museums",
    y = "Price",
    x = "") +
  scale_y_continuous(labels = scales::dollar) +
  coord_flip() +
  theme(legend.position="none")

```

