# Results: Vis for PB in Chicago

Anon

2024-03-19

## Load preprocessed data

In the file `process.py` we wrangled user log data from nested JSON into a table with one entry per PB project per user. Columns contain information on how each participant sorted and allocated funding to these projects

```
df = read_csv("processed_data.csv")
```

```
## Rows: 104 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): user, project
## dbl (6): sort_rank, pre_rank, post_rank, pre_allocations, post_allocations, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(df)
```

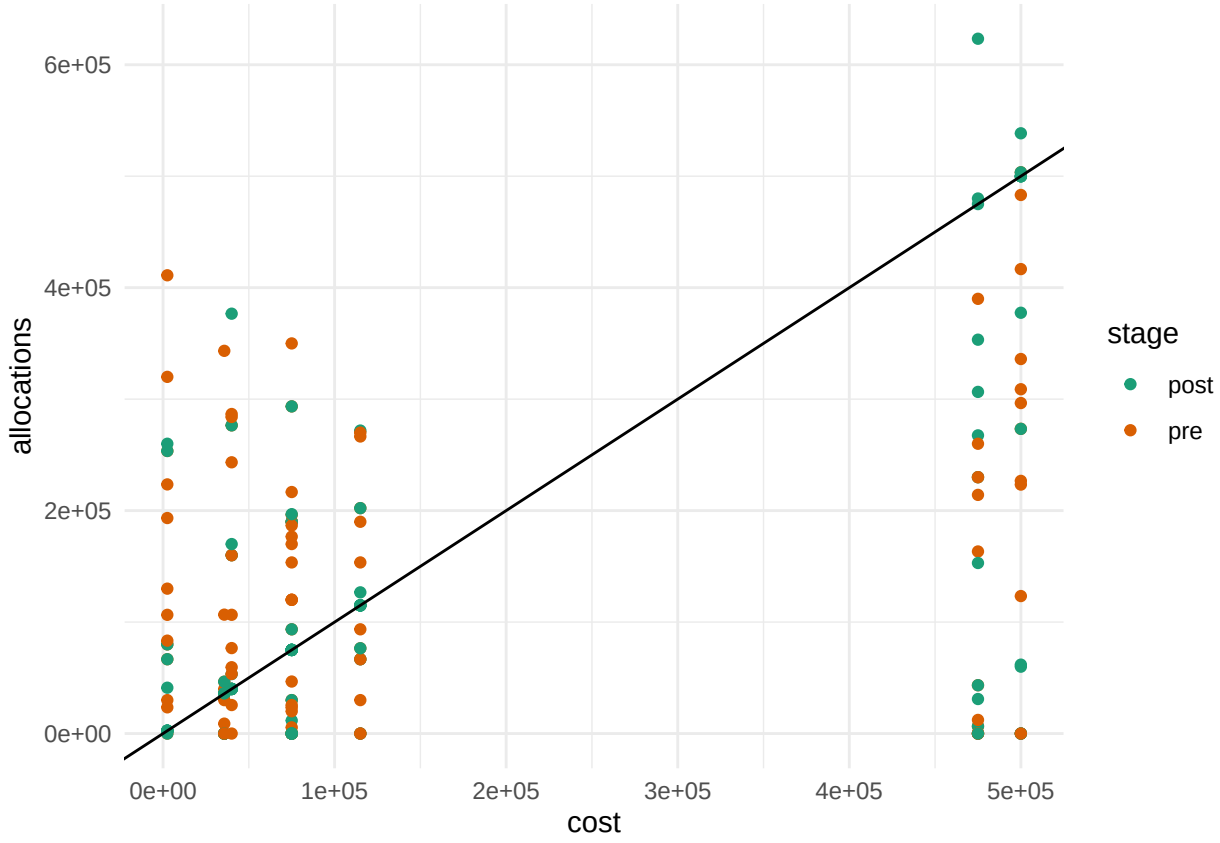
```
## # A tibble: 6 x 8
##   user  project    sort_rank pre_rank post_rank pre_allocations post_allocations
##   <chr> <chr>          <dbl>    <dbl>     <dbl>           <dbl>            <dbl>
## 1 p01   Bike Lanes         1        3         3          190000           190000
## 2 p01   School Im~         2        1         1          273333           273333
## 3 p01   Food Pant~         3        5         5           80000            80000
## 4 p01   Curb Cuts          4        4         4          160000           160000
## 5 p01   Street Li~         5        6         6           66667            66667
## 6 p01   Street Mu~         6        7         7               0                0
## # i 1 more variable: cost <dbl>
```

## Allocations

First, we examine how participants allocated dollars to projects before and after seeing what these projects cost.

```
df |>
  pivot_longer(
    cols=c('pre_allocations', 'post_allocations'),
    names_to = c("stage", ".value"),
    names_sep = "_"
  ) |>
  ggplot(aes(x = cost, y = allocations, color = stage)) +
  geom_point() +
  scale_color_brewer(palette = "Dark2") +
```

```
  geom_abline() +
  theme_minimal()
```

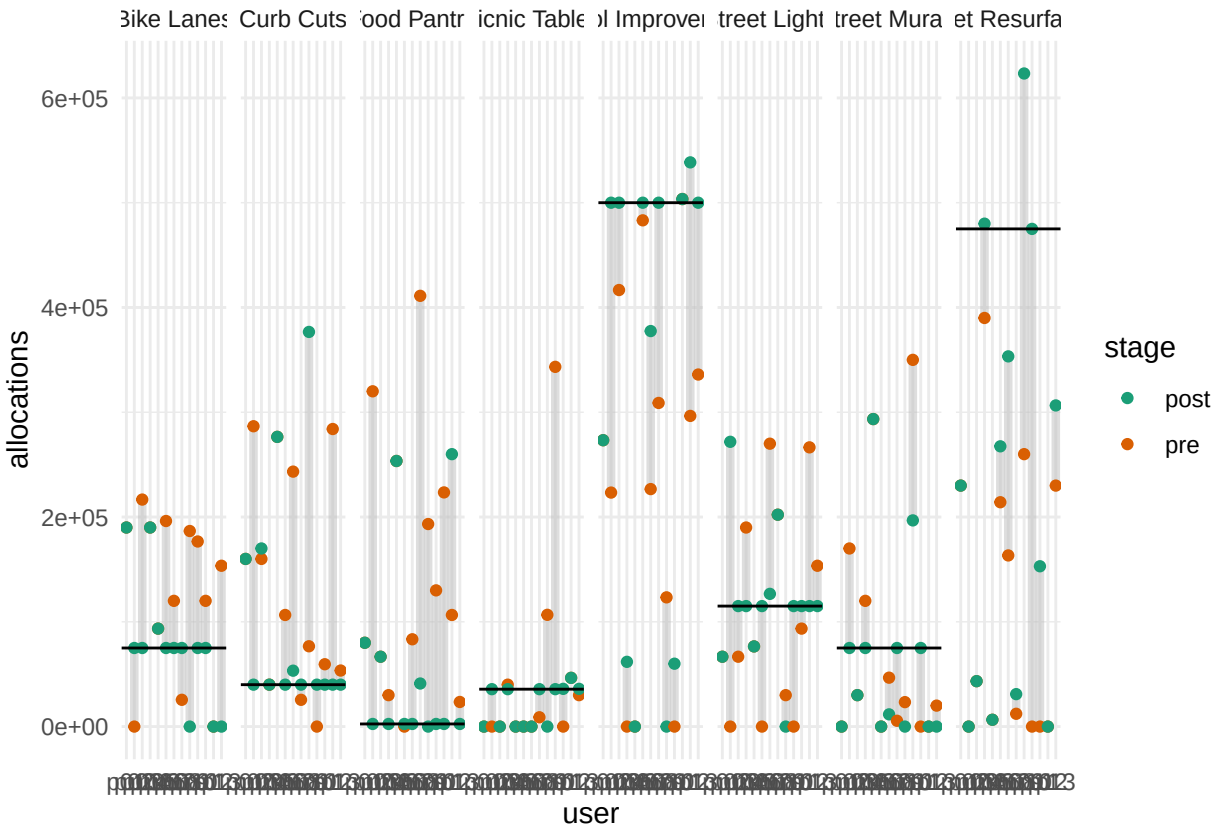



We can see some movement towards the estimated cost of projects after it's revealed for most projects (separated above at different levels of cost on the x-axis).

Let's create a view of these data that does a better job of showing the change per participant on each project.

```
df |>
  pivot_longer(
    cols=c('pre_allocations', 'post_allocations'),
    names_to = c("stage", ".value"),
    names_sep = "_"
  ) |>
  ggplot(aes(x = user, y = allocations, color = stage, group=user)) +
  geom_line(color="grey", size=1.5, alpha=0.5) +
  geom_point() +
  geom_hline(aes(yintercept = cost), color = "black") +
  scale_color_brewer(palette = "Dark2") +
  theme_minimal() +
  facet_grid(. ~ project)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




This gives a much better view of the heterogeneity in the pattern described above, where participants tend to follow the estimated cost. A few high-level notes on the chart above:

- Street resurfacing is an exception to the "set to cost" pattern.
- Participants seem particularly inclined to adjust allocations down for bike lanes, curb cuts, and food pantry.
- Participants seem to adjust up for school improvements.
- Sometimes participants don't update their allocations after seeing costs, which shows up as no visible pre-allocation or connecting line.

We think these patterns reflect the way that priorities get shuffled around once costs are known. Street resurfacing is the sacrificial project that people tend to under-fund, which poses a problem insofar as menu money is the only source of funding that Wards regularly receiving for road repairs. In contrast, participants are far less willing to compromise on schools. Other social priorities like bike lanes, curb cuts, and food pantry tend to receive less funding after costs are know than participants initially indicate they would like to spend on these projects. These numbers are not representative of an actual PB vote, but they show that even among experts in PB and urban planning, priorities can "go out the window" once costs are introduced to the display.

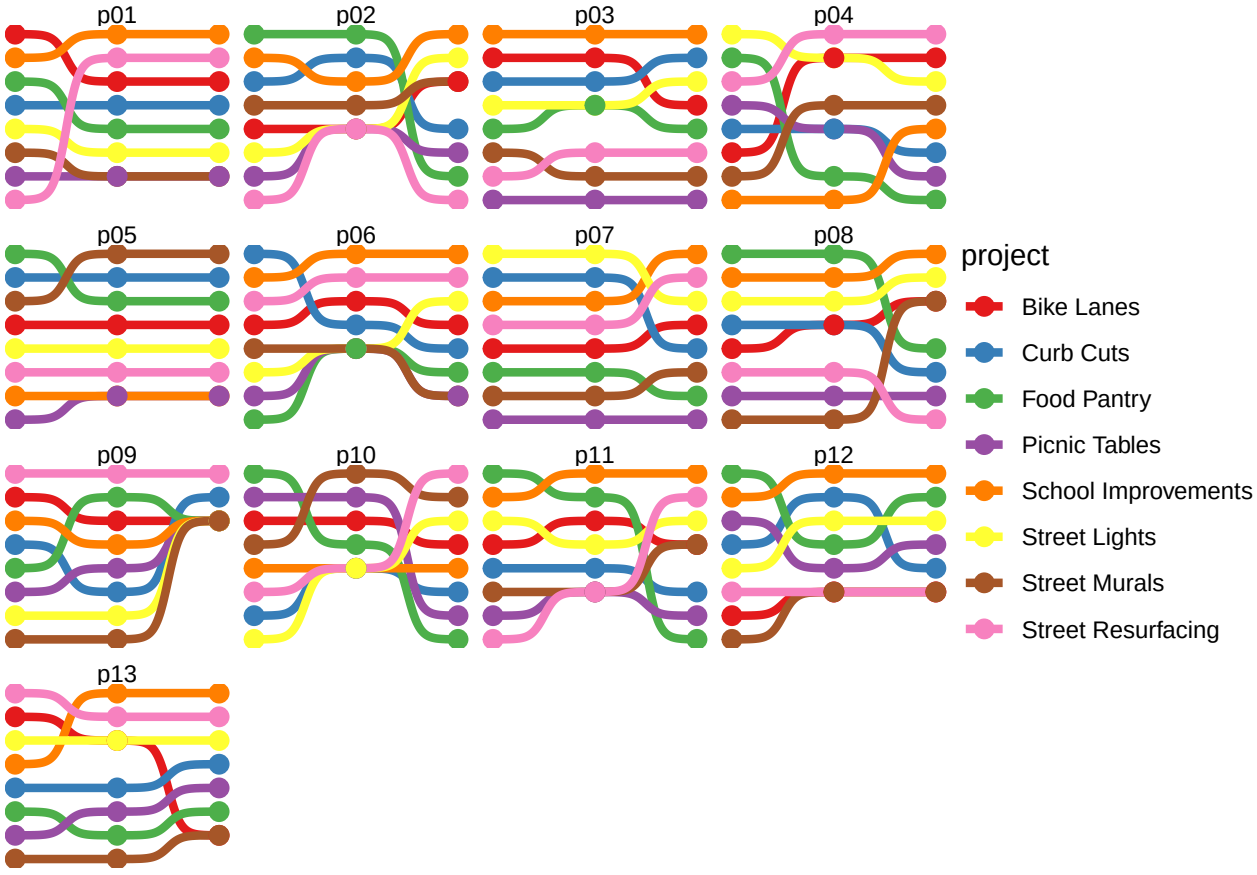
## Rankings

Next, let's look at how the ranking of projects that participants give changes across stages of graphical eliciation in our design probe. We look at three stages:

1. Sort order (sort): participants' initial sort of project by order of importance
2. Pre-allocations (pre): the order of projects implied by funding amounts before seeing costs
3. Post-allocations (post): the order implied by revised funding amounts after seeing costs

```
df |> pivot_longer(
    cols=c('sort_rank', 'pre_rank', 'post_rank'),
```

```
  names_to = c("stage", ".value"),
  names_sep = "_"
) |>
mutate(
  stage_n = case_when(
    stage == "sort" ~ 1,
    stage == "pre"  ~ 2,
    stage == "post" ~ 3
  )
) |>
ggplot(aes(x = stage_n, y = rank, color = project)) +
  geom_bump(size = 1.5) +
  geom_point(size = 3) +
  scale_color_brewer(palette = "Set1") +
  scale_y_continuous(trans = "reverse") +
  theme_void() +
  # theme(legend.position = "none") +
  facet_wrap(. ~ user)
```




We can see there's a fair amount of movement across stages of graphical elicitation, suggesting that these stages measure different notions of preference. A few high-level notes on the chart above:

- Curb cuts and food pantry tend to drop in rank between sort and pre.
- Street lights, murals, and school improvements tend to rise in rank from pre to post.
- Participants differ in their tendency to change project rankings between sort and pre vs between pre and post.

There's a lot of heterogeneity here, and we don't want to over-interpret small-N results. However, it's worth noting a few dynamics that may help to explain these rank changes. Inexpensive projects with high social impact (e.g., curb cuts and food pantry) tend to go down in rank between sort and pre, suggesting that dollars are the wrong unit of value for these priorities. On the other hand, inexpensive projects that may not initially seem as important (e.g., street lights and murals) tend to go up in rank between pre and post. We speculate this is because participants realize at this stage that they have enough money to fully fund these projects, even if they cannot fully fund other priorities. In other words, there may be a bias toward fully funding as many projects as possible, which is not ideal from the perspective of trying to measure preferences.

## Summary

The main take-away from our log data is that reasoning in terms of costs changes apparent preferences, as does reasoning about dollars. We caution that these results are from 13 expert participants and are not representative of the general population that would use a PB interface. Still, these patterns are of interest for tool-builders, researchers, and policy workers thinking about how to scaffold preference elicitation.