Patterns in the Usage of Homeless Services

An Exploratory Analysis for Community Technology Alliance

Sangyoon Park and Kris Sankaran

May 5, 2016

1 Introduction

This document describes an exploratory analysis of data provided by Community Technology Alliance (CTA), an organization seeking to improve the use of data in the design and provision of homelessness services. Its authors are volunteers associated with DataKind and Stanford Statistics for Social Good. Our purpose in writing this report is to follow up the approaches developed during a November 2015 DataKind DataDive and set the stage for in-depth study. The data includes records from Monterey and Santa Cruz Counties collected between 2010 and 2015.

Much of our focus here is in recording intutition we have gained about the data so far. Since our goal in this preliminary pass is to better understand the data from a coarse level, we have intentionally passed over some of the specific features to which we have access. Our justification is that it can be very frustrating to try to place details into their proper context without a higher-level view of the data.

We imagine this report being a relevant reference for future Data Ambassadors or Data Corps members working with CTA, even if the questions they are interested in are more specific than the ones we study here – we hope this document can help transfer some of our intuition, which can then be incorporated into different analysis. For this reason, we have not filtered down to only the transitional housing projects (which had been the focus of the DataDive), as we would like this reference to be useful even if the Data Corps tackles a problem unrelated to transitional housing.

We have taken care to ensure the findings in this report are reproducible. All preprocessing is performed in a self-contained script, described briefly in Section

3, and this report is actually written in Rmarkdown, so all figures can be generated by running a single script. All of this code is available in a private repository[1]. Finally, we note that this report only includes those figures that we found especially interesting; many variations on the ideas presented here can be found in our original code repository.

Finally, note that this exercise is completely exploratory. While we appreciate rigorous statistical modeling and inference, we suspect this formal work can be done better after an initial informal exploration. We welcome your feedback as we scope out the next stages of our analysis.

2 Raw Data Views

As a first step towards becoming acquainted with the raw data, we generated univariate histograms for all available features. Many of these lent themselves to natural interpretation – see for example Figures 1, 2, 4. Figure 1 shows that the ages of clients served across the data are multimodal, and vary between those who are and are not chronically homeless (interestingly, there are a number of chronically homeless clients younger than 25). Figure 4 examines the distribution of the number of times each client participates in each type of project. We can see that many participants join shelter projects multiple times while other projects rarely have more than two re-entries.

However, some of the histograms were more difficult to glean useful information from. Consider Figure 3 for example – we can only gather any useful information from this feature after filtering to the top 25 most common services. These kinds of situations motivated the preprocessing explained in Section 3.

3 Preprocessing

As observed during the November 2015 DataDive, working with this raw data is difficult for several reasons,

- Some fields have many levels, which are often related to one other. For example, the levels `Rental by client, with VASH subsidy` and `Rental by client, with other ongoing housing subsidy` in the `Destination` field of the `raw_EntryExit2010-1001--2015-0930.csv.xls` might perhaps be combined. Table 2 shows the original entry / exit names, along with the group that we associated them with, and Table 3 gives the original project types, and the groups to which we assigned them.

[1] But don't hesitate to request access.

Term	**Definition**
Entry-exit pair	The interaction described by the `Project Entry ID` in `raw_EntryExit2010-1001--2015-0930.csv.xls`. This can be interpreted as a client's participation in a homeless service program. During each interaction, an entry and exit condition is recorded, as specified by the fields `Living situation before program entry?` and `Destination`. We typically refer to the entry-exit types, rather than the detailed entry-exit values. Each client in the data may be involved in multiple programs, possibly at different service organizations, and possibly in overlapping time frames.
Entry-exit type	A coarser view of the `Living situation before program entry?` and `Destination` columns, determined during preprocessing. For the mapping between original and grouped entry-exit values, see Table 2.
Project / Program	The value of the `Project ID` field, specified in `raw_Project2010-1001--2015-0930.csv.xls`. These are interpreted as individual programs run by larger organizations.
Project type	A grouping of the projects / programs according to the type of service they provide. This is a feature derived from the field `HUD Standards Information Program Type Code` in `raw_Project2010-1001--2015-0930.csv.xls`, the formal mapping is given in Table 3.

Table 1: A formal definition of the terms used throughout this report.

Figure 1: Age histogram, across clients who are not and are chronically homeless (top and bottom panels, respectively).

Figure 2: Breakdown of entry-exit pairs by entry and exit types, faceted across project types.

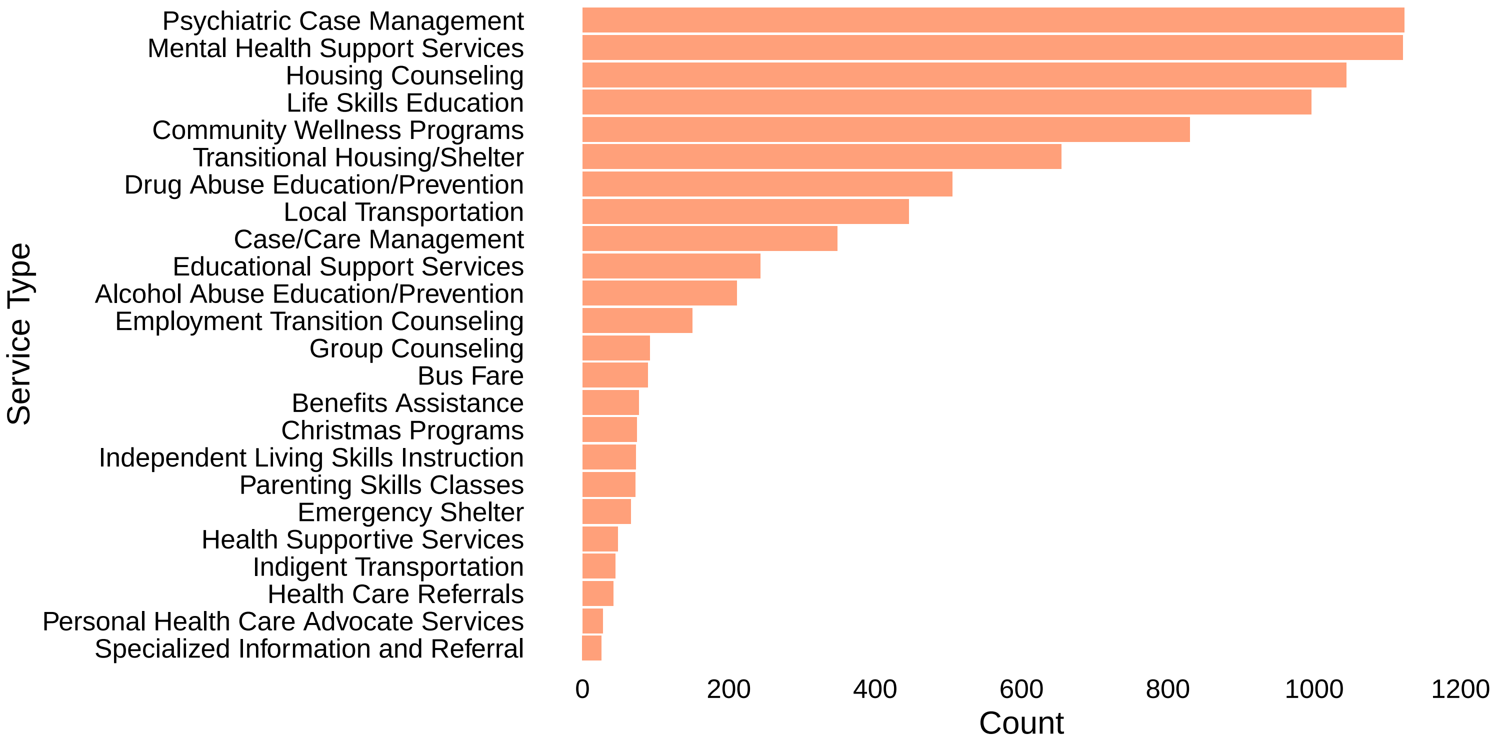

Figure 3: Top 25 most frequent services within transitional housing projects, out of 55 total.

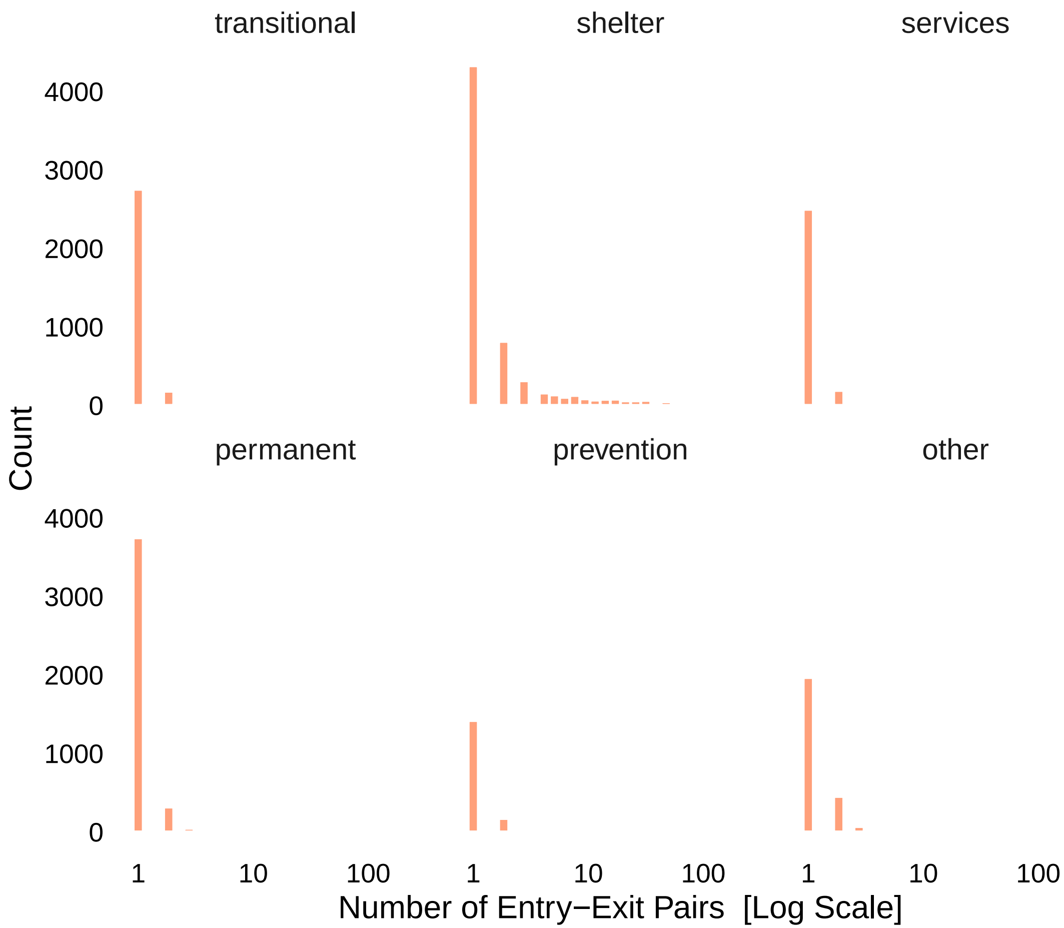

Figure 4: Histogram of the number of entry-exit pairs for each client, faceted across project types.

- As seen in the previous bullet, many of the levels have long names. While it is good to be descriptive in record keeping, such naming conventions can be unwieldy to use in computational work.
- Many features have levels with very low counts. For example, of the 17010 rows of the `raw_client2010-1001--2015-0930.csv.xls` table, only 23 have the categorical level equal to `Data not collected (HUD)`.
- Some features have many NAs. For example, in `raw_EntryExit2010-1001--2015-0930.csv.xls`, 32279 of the 51890 available rows are NA.

Some of these points are related to the idea that, when working with highly annotated data, coarse level inference becomes hard. For example, when rentals are divided into many types, it is hard to see what is happening with rentals in general. Other points are just related to ease of computational use – it is hard to make plots when names appearing on the screen are 40 characters long. To get around these issues, we made the following preprocessing decisions,

- For features with many levels (or levels with low counts) we try to define new features which group similar levels together.
- We shorten level names.
- We omit features with too many NAs.

To see the benefit of grouping levels into project and entry-exit types, consider Figure 2, which describes the proportional breakdown of entry-exit types within each project category, where columns and rows correspond to entry and exit types, respectively. If we were to use the original entry-exit variables, each cell of this figure would be a 31×31 matrix, each with relatively few samples. With this coarser view, we can make several meaningful observations:

- For permanent, prevention, and services projects, participants mostly enter and exit with the rental status.
- From shelter projects, participants mostly transition 1) from homeless to service, or 2) from service to service.
- Participants in transitional projects mostly enter from service or social and exit as rental or social.

Next, as observed during the DataDive, joining the (8) tables in a meaningful way can be problematic. We have avoided trying to build a single master table, and instead join fields on an as-needed basis.

During preprocessing, we identified several minor anomalies in the data. We have written about them in the Appendix, in case future data analysts are interested, but our work was not affected in any way. The only anomaly that might be worth special follow-up is the fact that entry and exit incomes are almost always equal to each other, as income is a potentially important feature for future analysis.

All preprocessing steps are contained in `src/processing/processing.R` in our code repository. This script takes the 8 raw csv files and creates 8 processed versions, which are stored in our `data/processed_data` folder on Google Drive.

4 Entry-Exit Macro Views

We care about the entry-exit table because it links clients and the programs in which they participate – other tables give either client or project specific information. In light of this, we guided our initial exploration towards developing a sense of entry-exit patterns across client and project features. Towards this, we first sought a "macro" view of the entire landscape of homelessness service, rather than focusing on individual clients. A client-level examination is discussed in Section 5.

In this section, we provide a sequence of figures that trace this exploration, giving commentary on what we see in each figure. We have intentionally tried to pack as much information into each visualization as possible, and we hope that you can discover details that we might have not yet commented on. To paraphrase Edward Tufte in *Envisioning Information*, it is a good sign when a visualization rewards careful study.

4.1 Full Entry-Exit Data

Our first step is to display the raw data points with entry and exit types within the 7 (out of 11) categories we defined during preprocessing; this preserves about 80% of the original entry-exit pairs and results in Figure 5. The entry and exit types we kept are indicated in the figure legend.

We start by describing how to read this display. Each vertical line represents a single entry-exit event, from start to finish. The vertical axis indexes time; the horizontal axis indexes clients. The colors at the line endpoints indicate the types of entry and exit. We know that this plot is difficult to read, and the rest of this section is focused on different refinements. Nonetheless, we can identify a few takeaways.

First, we see a long "tail" along the bottom left. The data seems to have been generated by filtering down to individuals who exited after 2010 (see the Appendix for more on this point), but there are a few clients who entered long before 2010 – some as early as 2004! In fact, this tail obscures some of the detail where most entry-exit pairs happen, so we regenerate the figure, keeping only clients who entered after 2010. The zoomed view is in Figure 6.

```
##                                                                            Destination_label destination_group
##  1:                                                   Place not meant for habitation (HUD)          homeless
##  2:                                  Staying or living with friends, permanent tenure (HUD)            social
##  3:                                                                            Other (HUD)             other
##  4: Emergency shelter, including hotel or motel paid for with emergency shelter voucher (HUD)        service
##  5:                                                                                     NA                NA
##  6:                                   Staying or living with family, permanent tenure (HUD)            social
##  7:                                      Rental by client, no ongoing housing subsidy (HUD)            rental
##  8:    Staying or living with family, temporary tenure (e.g., room, apartment or house)(HUD)           social
##  9:                                                              Client doesn't know (HUD)             other
## 10:                                                      No exit interview completed (HUD)     not_collected
## 11:                          Hospital or other residential non-psychiatric medical facility (HUD)      medical
## 12:    Staying or living with friends, temporary tenure (e.g., room apartment or house)(HUD)           social
## 13:                                 Substance abuse treatment facility or detox center (HUD)           medical
## 14:                                 Rental by client, with other ongoing housing subsidy (HUD)          rental
## 15:                                  Psychiatric hospital or other psychiatric facility (HUD)          medical
## 16:                  Transitional housing for homeless persons (including homeless youth) (HUD)        service
## 17:                                   Permanent housing for formerly homeless persons (HUD)              owner
## 18:                                                 Rental by client, with VASH subsidy (HUD)           rental
## 19:                                                              Data not collected (HUD)     not_collected
## 20:                          Hotel or motel paid for without emergency shelter voucher (HUD)            hotel
## 21:                                       Jail, prison or juvenile detention facility (HUD)       institution
## 22:                                                                  Client refused (HUD)             other
## 23:                          Residential project or halfway house with no homeless criteria (HUD)      service
## 24:                                                                        Deceased (HUD)          deceased
## 25:                                                                     Safe Haven (HUD)           service
## 26:                                     Owned by client, with ongoing housing subsidy (HUD)             owner
## 27:                                       Owned by client, no ongoing housing subsidy (HUD)             owner
## 28:                                                                         Homeownership             owner
## 29:                                        Foster care home or foster care group home (HUD)       institution
## 30:                                           Long-term care facility or nursing home (HUD)       institution
## 31:                                                            Rental room/house/apartment            rental
##                                                                            Destination_label destination_group
```

Table 2: Mapping from original entry-exit types to our grouping.

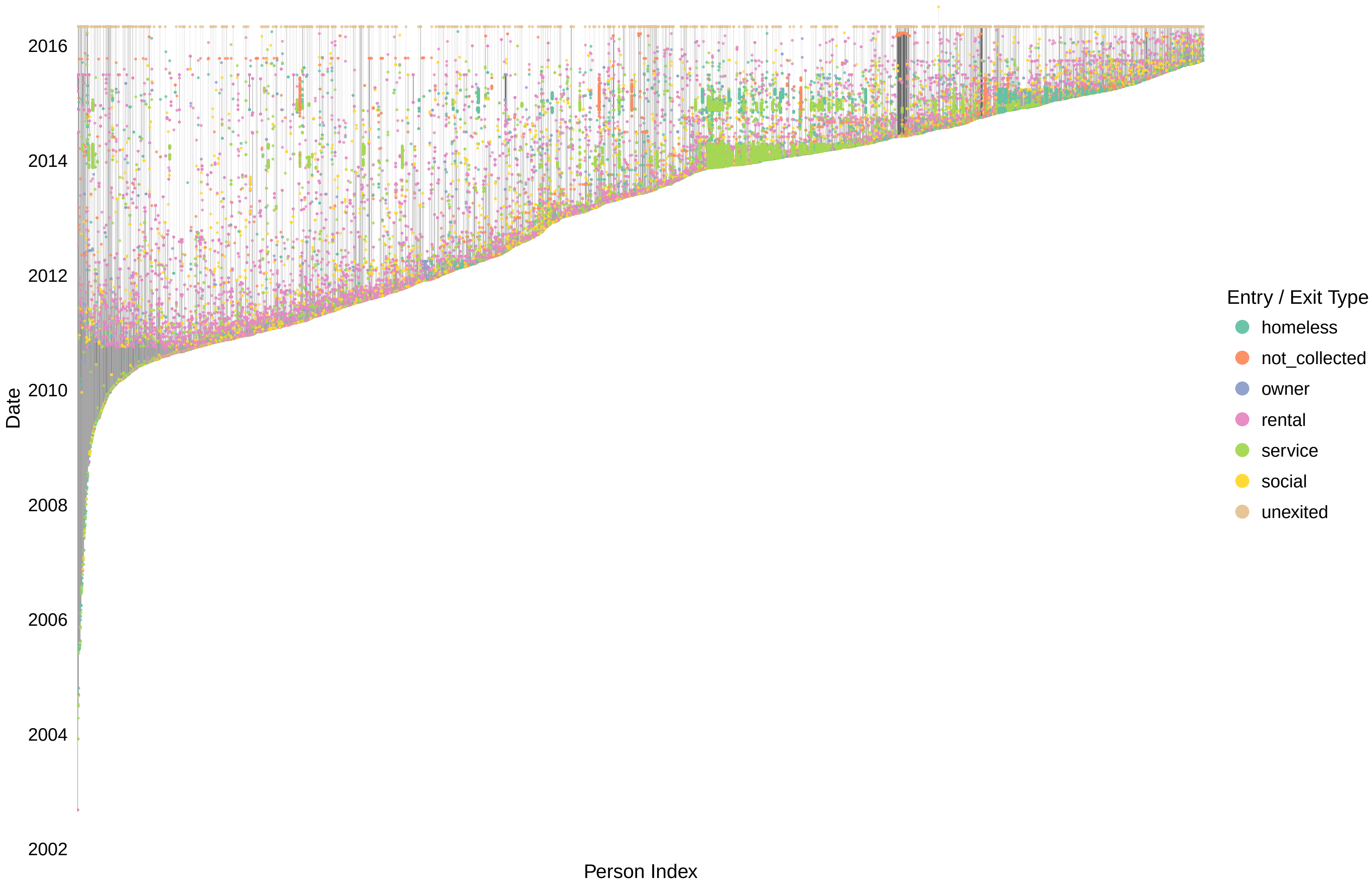

Figure 5: Entry-exit pairs, plotted across time.

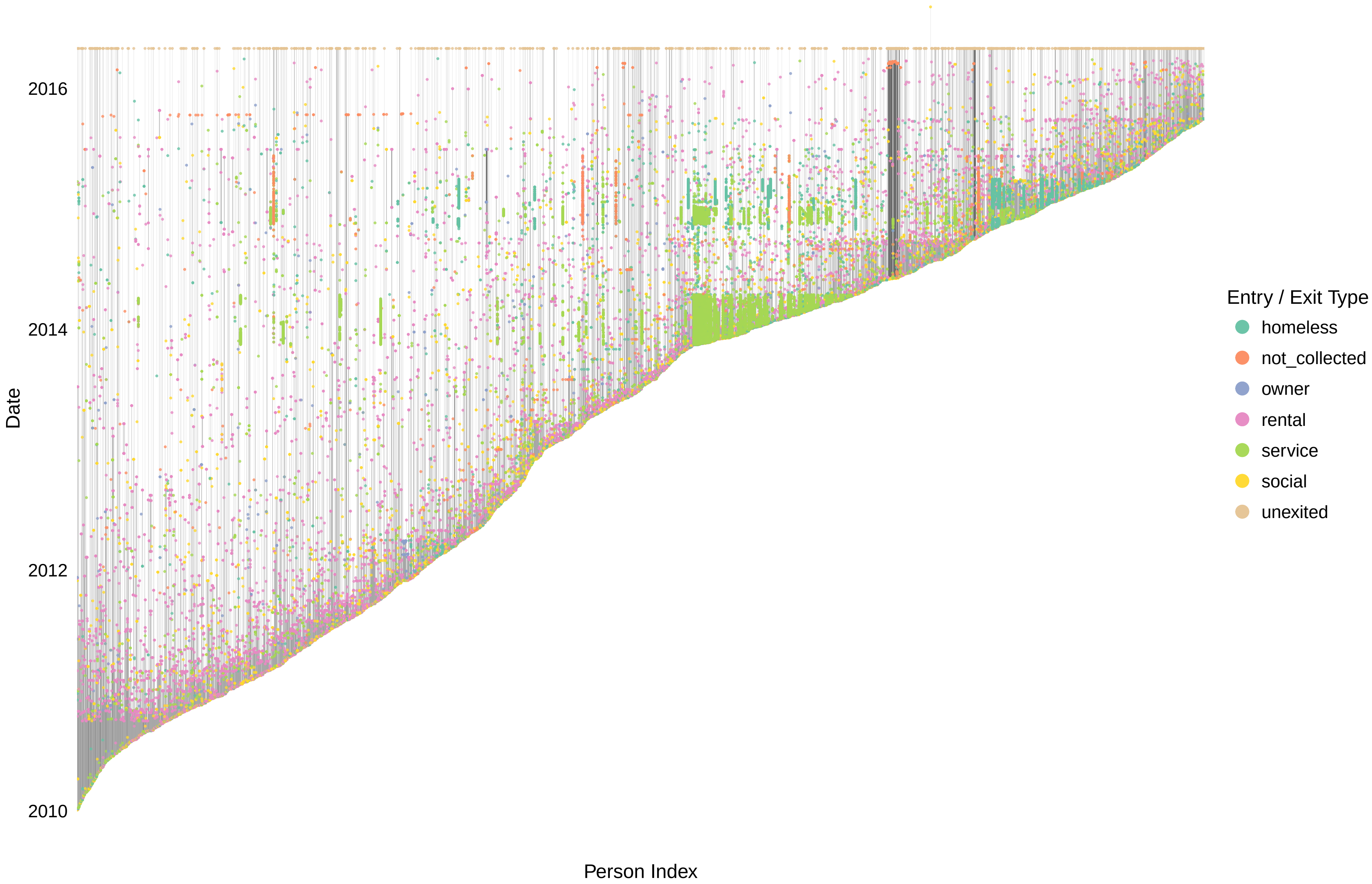

Figure 6: Entry-exit pairs, filtered to entries after 2010.

```
##                                                                      hud_copy HUD Standards Information Program Type Code
## 1:                                                              Other (HUD)                                          other
## 2: PH - Permanent Supportive Housing (disability required for entry) (HUD)                                      permanent
## 3:                                                  Emergency Shelter (HUD)                                        shelter
## 4:                                                      Services Only (HUD)                                       services
## 5:                                               Transitional housing (HUD)                                   transitional
## 6:                                            Homelessness Prevention (HUD)                                     prevention
## 7:                                                    Street Outreach (HUD)                                          other
## 8:                                                            RETIRED (HUD)                                      permanent
## 9:                                              PH - Rapid Re-Housing (HUD)                                      permanent
```

Table 3: Mapping from original project types to our grouping.

Next, we can see general changes in the shape of the lower envelope traced out by the data points. This represents temporal variation in the total number of clients who are entering the system. For example, a slight slowdown in the number of entries around 2012 can be seen as the envelope becoming slightly more vertical; we also notice a big increase in the number of entries around mid-to-late 2013, where the envelope becomes more horizontal.

We also see a few long vertical strings of green and blue points. These occur when the same person makes many homeless or service entries and exits within a very short time span. On a similar note, there seems to be a large "triangle" of green and blue points in late 2013 and early/late 2014, respectively. These may represent specific organizations providing long strings of consecutive entry-exits to their clients during this period. Similarly, the horizontal bands visible in a few regions in the display represent dates where many clients exited at exactly the same time. This is probably due to an organization scheduling a mass exit (or at least recording exits that way in their data system). This kind of behavior was considered during the DataJam, but it seems more common than we had anticipated then (we had focused more on a single date).

Next, we interpret the density of points across the display. The density is highest around the bottom of lower envelope. Indeed, if every individual only ever interacted with the system, for at most one year, there would be an upper envelope exactly one year higher than the lower envelope. That this is not the case means that clients are reentering the system. Nonetheless, the density is quite a bit lower in this upper-left reentry region compared to the first entry region near the lower-edge of the envelope (say, ascending about 3 months over the lower envelope). So, most clients have entry-exit lengths of about four months, and never re-enter, though a reasonable fraction do.

Consider the upper envelope of beige points. These represent clients who have not exited the system yet[2]. While the number of clients who have not exited

[2]Formally, they have NA values in the exit date of the entry-exit table.

yet is very high among recent entries (2015), which are near the top right of the plot, there are actually quite a few clients who had many years ago but have not yet exited. These clients are more numerous than we initially expected. Also, notice one client with a future exit date!

4.2 Faceted by Entry-exit Type

It is hard to compare entry and exit types (colored dots) when they are all overlaid. Indeed, it is not even clear from this figure what the most common entry and exit types are. So, we decided to split it by entry and exit types, and the result is shown in Figure 7. It can be read in exactly the same way as the previous display, except each cell contains entry-exit pairs of a specific configuration. The rows and columns correspond entry and exit types, respectively. For example, the top-left cell includes clients who both entered and exited as homeless; the bottom-right cell includes clients who entered social but have not exited yet.

The rental and service columns seem to have most of the points, so these are the most common exit types. However, the structure of the entry-exit pairs between these columns is quite different – rental exits tend to have long stays in the system, while service exits are usually out quite quickly. On this note, consider the rental entries. Almost everyone who enters rental also exits rental.

Let's study the homeless entries and exits more carefully. Consider the homeless exits (the first column). Most points are in the top row, which means that most homeless exits are from clients who started homeless, which makes sense. However, a number of those who exit homeless actually started in the service category. Now, consider the top row – these are clients who enter homeless. Among these homeless entries, a large number exit to the rental and service categories.

Note that there are clients entering the system as owners; this is the "owner" row in the table. They are rare, but they are present, and they do not always even exit as owners.

Recall the interpretation from the previous section about the shape of the lower envelope and the density of the re-entry region. We notice that these characteristics are not uniform across entry-exit combinations. For example, the number of social, rental, and owner exits is relatively stable across time, while the number of homeless and service entries have a few points of rapid acceleration, presumably due to the creation of new organizations (or their introduction into CTA's data collection system). The increasingly horizontal shape of lower-envelope in the "unexited" column makes sense – there are more unexited entry-exit pairs as we get closer to the present.

We note a limitation of this display, compared to the previous one. Here, we have lost continuity in each individual; the same person can be appearing in several entry-exit pair types, just at different times. That is, they may appear across several cells in the grid, explaining why the lower envelope is no longer continuous,

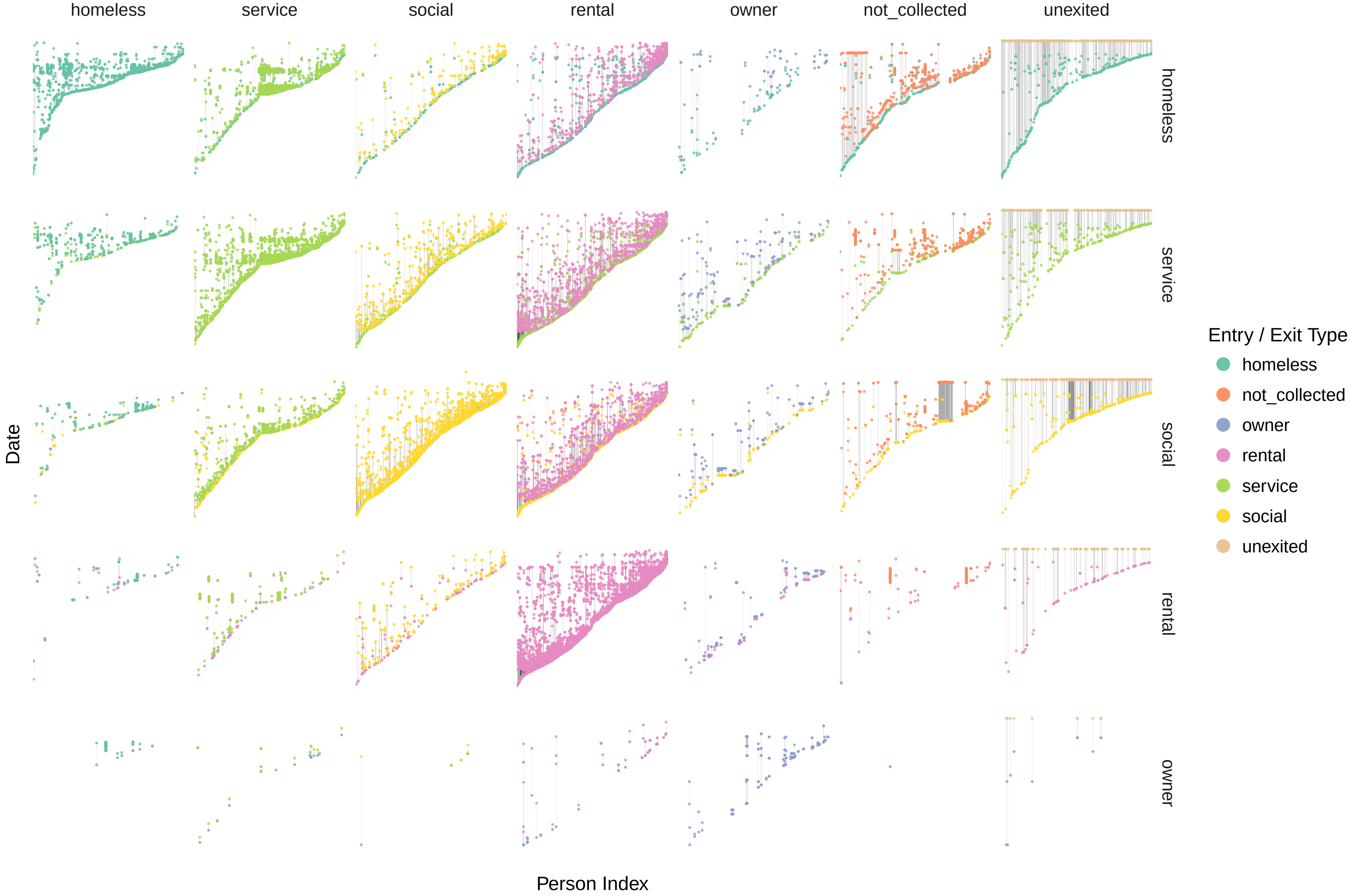

Figure 7: Entry-exit pairs entered after 2010, faceted across entry and exit types.

4.3 Faceted by Entry and Project Type

After investigating variation in entry-exit pairs across client characteristics, we became interested in how this variation might appear at the project level. Indeed, some of the most dramatic differences in the previous plot were between rental and service entry-exit pairs, which we imagine being associated with the type of programs in which clients enroll. To test this assumption, we created Figure 8.

This plot like the others, except now the columns split entry-exit pairs by project type rather than exit type.

Now, some comments. Consider the "permanent" project type in the second column. There appear to have been more projects of this type earlier in the data, which is followed by a dry spell and then by renewed activity in the last few years. Furthermore, in the earlier period, the permanent housing projects catered mostly to entries from the rental category, while now there are more entering from the homeless and service groups. Further, among the homeless entries in this earlier period, most seem to have never exited (see the beige points at the top). In comparison, almost all the rental entries from this period did exit.

Another feature that stands out is the shape of the "prevention" project type. There seems to have been a prevention organization that operated briefly and had a mass exit. More recently though, there have been more entries in prevention again, but catering more to rental and social (rather than homeless) entries.

Next, recall the interpretation of the lower envelope and reentry region density. The transitional projects tend to have thicker and lower envelopes, meaning longer entry-exit pairs, but relatively fewer reentries (though there are a few that stand out). The "shelter", "services", and "other" categories seem to have tighter lower bands, but more frequent re-entries. We also notice a parallel lower band (colored in red, meaning exit type was not_collected) in the "homeless" entry-"services" project cell. We have no explanation for this.

Also, notice that, in the transitional projects column, the cells for social and service are generally much denser – this reflects the (relatively obvious) fact that most entries to transitional housing programs are either already involved with a service or are living with a friend. Also, notice that the lower envelope seems to be becoming somewhat more vertical in recent years, across all entry types. This means the number of clients entering transitional housing is decreasing slightly.

4.4 Filtering to Transitional Housing

Though we did not want to focus entirely on transitional housing projects, as in the DataDive, we thought it would still be valuable to consider entry-exits in the transitional project type alone. This is displayed in Figure 9. Remember that rows correspond to entry types, while columns describe exit types.

Our first reaction to this figure was that there actually have not been too many clients enrolled in transitional housing programs, at least in the data so far. It seems possible to examine each individual one at a time, in fact.

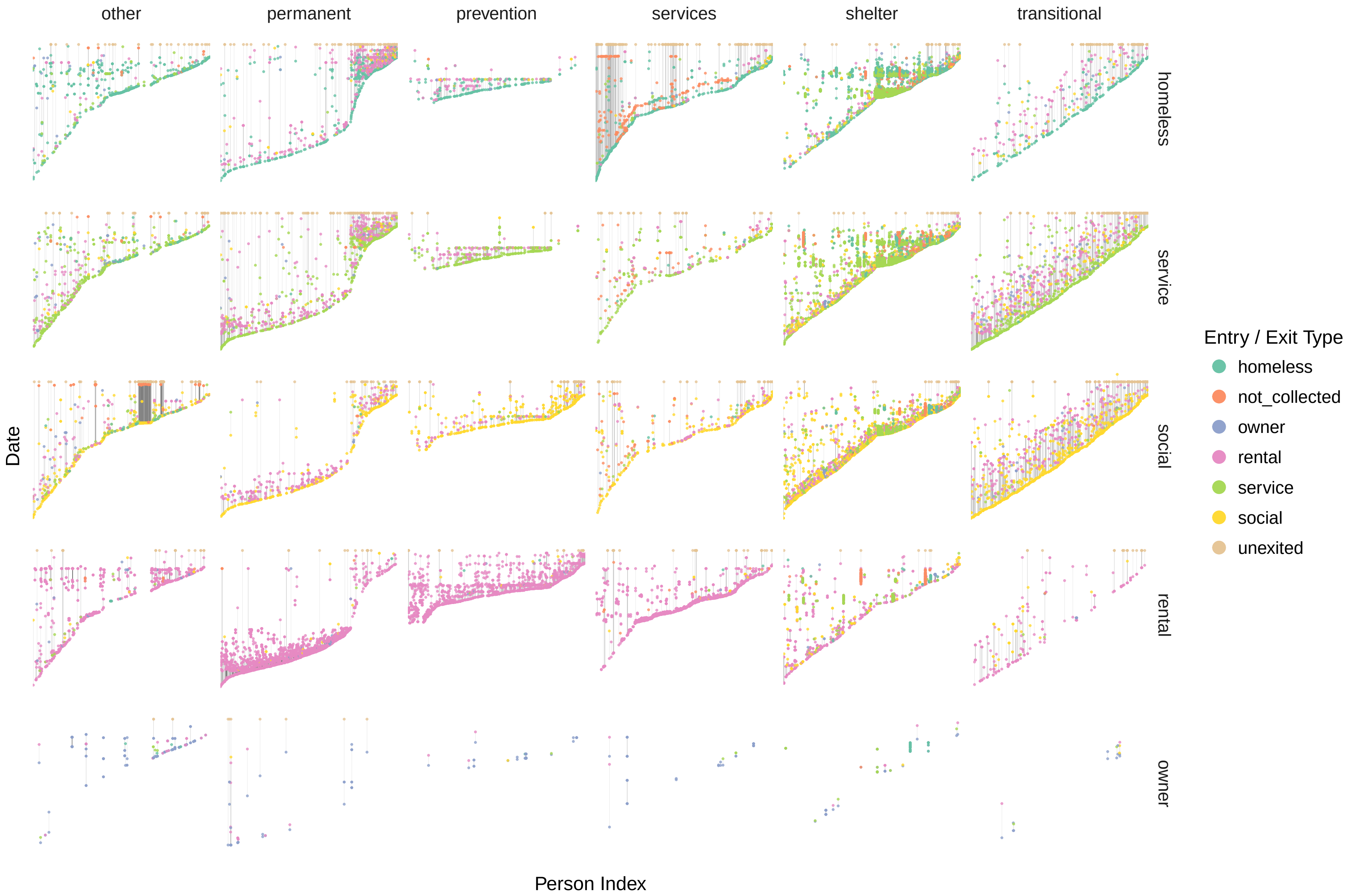

Figure 8: Entry-exit pairs entered after 2010, faceted across entry and project types.

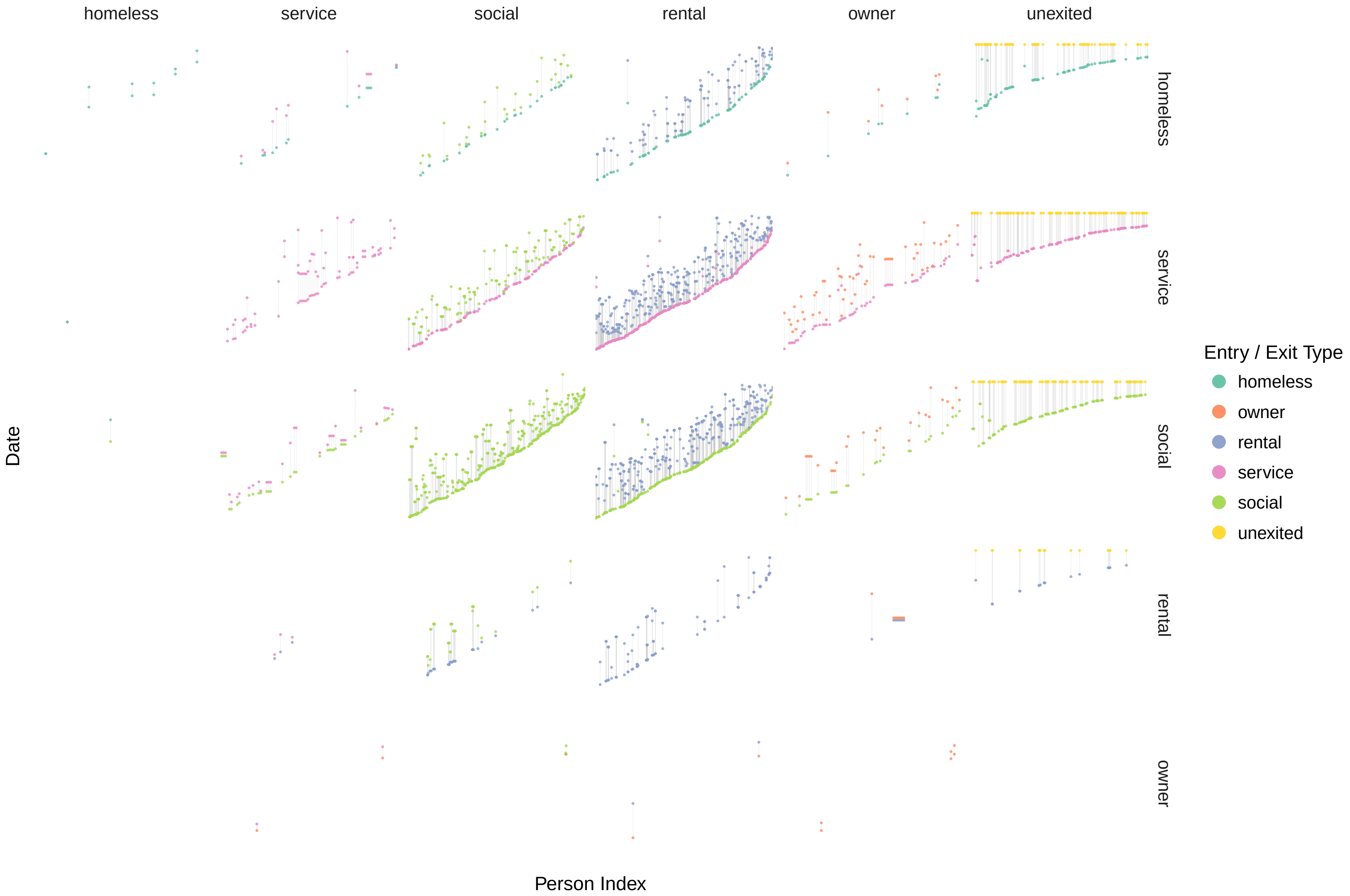

Figure 9: Entry-exit pairs entered after 2010, filtered to transitional housing projects.

In this view, we again see the fact that most entries to transitional housing programs come from social or service entries. We now additionally see that most exits are to rental.

We note that the shapes are relatively similar across rows, suggesting that, for transitional housing programs, exit types cannot be easily predicted from entry type alone. The possible exception is that homeless to social entry-exits seem to be somewhat less frequent (and shorter) than social exits among service or social entries, though the strength of this association does not seem *that* strong.

4.5 Faceting by Entry and Family Status

One of the notable findings of the November 2015 DataDive was that the length of entry-exits was quite different between families and non-families. To follow-up this idea, we modified the plot to display family information. Figure 10 displays family status along columns and entry type along rows.

We notice that those in families are less likely to reenter the system, reflected by the relatively lower density in the upper left region, which we suspect is not simply due to the smaller overall number of entry-exit pairs among families. We also see that families are historically less likely to enter homeless, though there seems to have been a burst of homeless family entries in the last year. Similarly, there seem to have been more rental family entries earlier in the data.

Also, notice that there has been an acceleration in the number of non-family entries, compared to family entries, in the last few years, which we can tell from the shape to the lower envelopes between the two columns.

4.6 Final Remarks on Entry-Exit Macro Plots

We have shared the macro views of the entry-exit data that we found most interesting. However, there are many other variables that we could have faceted by and many specific temporal windows we could have zoomed into, each can tell a slightly different story. In particular, by zooming into specific time periods, we can turn a "macro" perspective of the data into a "micro" view. The main downside of this approach is that comparisons across time become difficult, and there are many more plots to inspect. Nonetheless, we note that these plots have been generated in the `src/exploratory/person_level_plots.Rmd` document in the code repository.

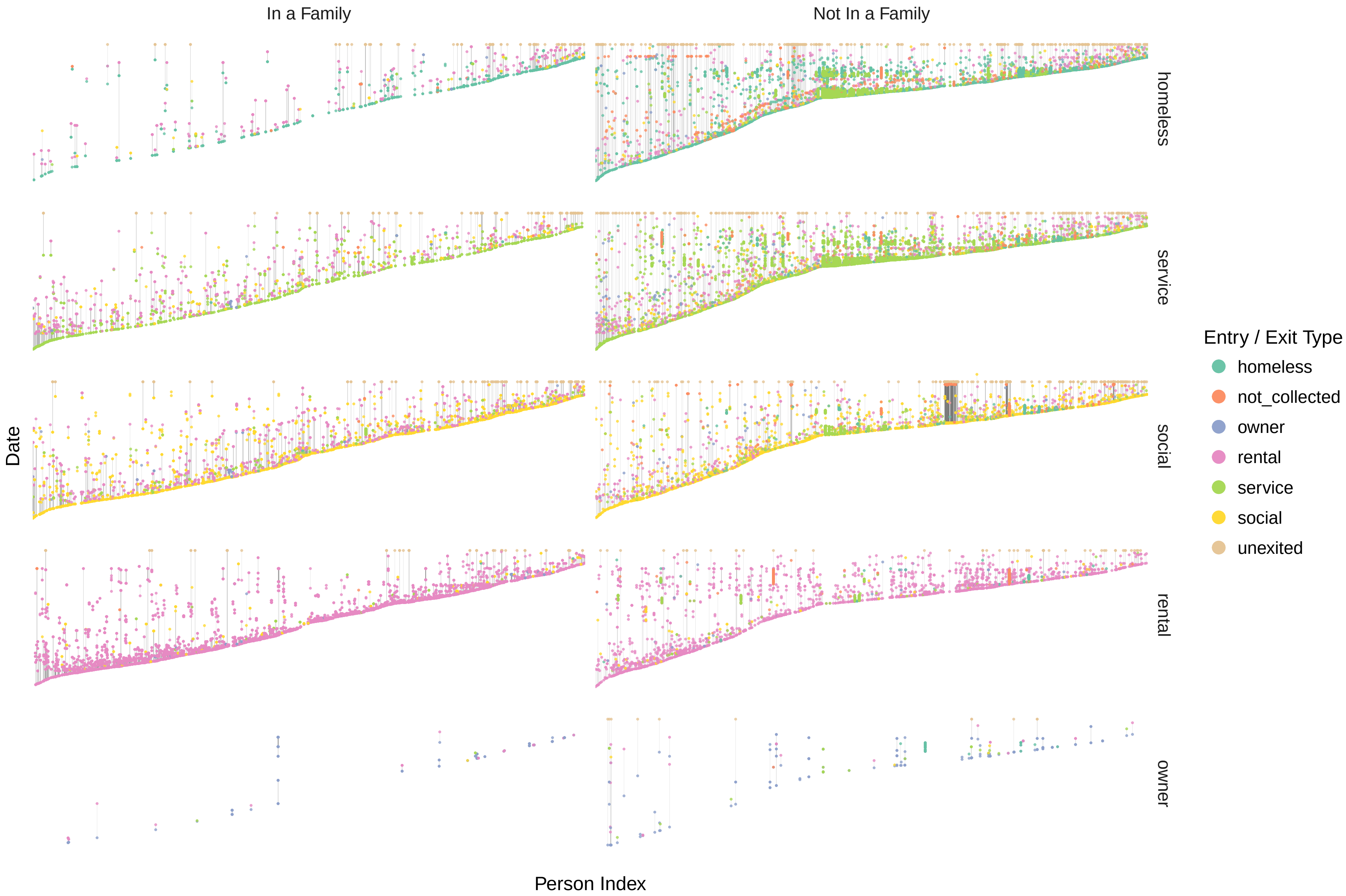

Figure 10: Entry-exit pairs entered after 2010, faceted by family status.

5 Client-Level Micro Views

To complement the "macro" displays developed in the previous section, we explored clients' interactions with the system at the individual level. Specifically, we plotted and examined each client's project participation timeline. This seems especially relevant for any individual-level modeling or inference we may study in the future.

We chose to focus on clients with 6 entry-exit pairs as there are enough clients to examine (more than 100 clients with 6 entry-exit pairs) and as we thought 6 entry-exits are numerous enough to reveal interesting patterns in project participation. It need be noted, however, that only 10% of clients have six or more entry-exit pairs. In fact, nearly 70% of clients in the data set have participated in only 1 project (cf. Figure 4).

After plotting timelines for all clients with 6 entries/exits, we could group them into three categories:

- Clients with consecutive participation in the same project, see Figure 11.
- Those with consecutive participation in different projects, see Figure 12.
- Those with other, less straightforwards patterns, see Figure 13.

Interestingly, we found that most entry-exits in the first category (consecutive participation in the same project) are for shelter projects. This strongly suggests that if the same project type is consecutively participated then it is very likely to be 'shelter'. It is also noteworthy that the periods over which these consecutive entry-exits take place is relatively short in most cases. From these observations, we deduce that multiple engagements within a short time can be used to identify shelter projects. It is also noteworthy that entry and exit types between these consecutive entry-exit pairs are often inconsistent, despite belonging to the same project type.

It is also interesting that "transitional" projects tend to co-occur or follow "other" projects (Figures 12 and 13). This may suggest that "transitional" projects often require some preliminary or complementary projects.

Of course, this very small sampling does not allow us to make claims about clients' project participation patterns in general. However, they suggest some patterns of client-system interaction are potentially worth following up or modeling in a more formal way.

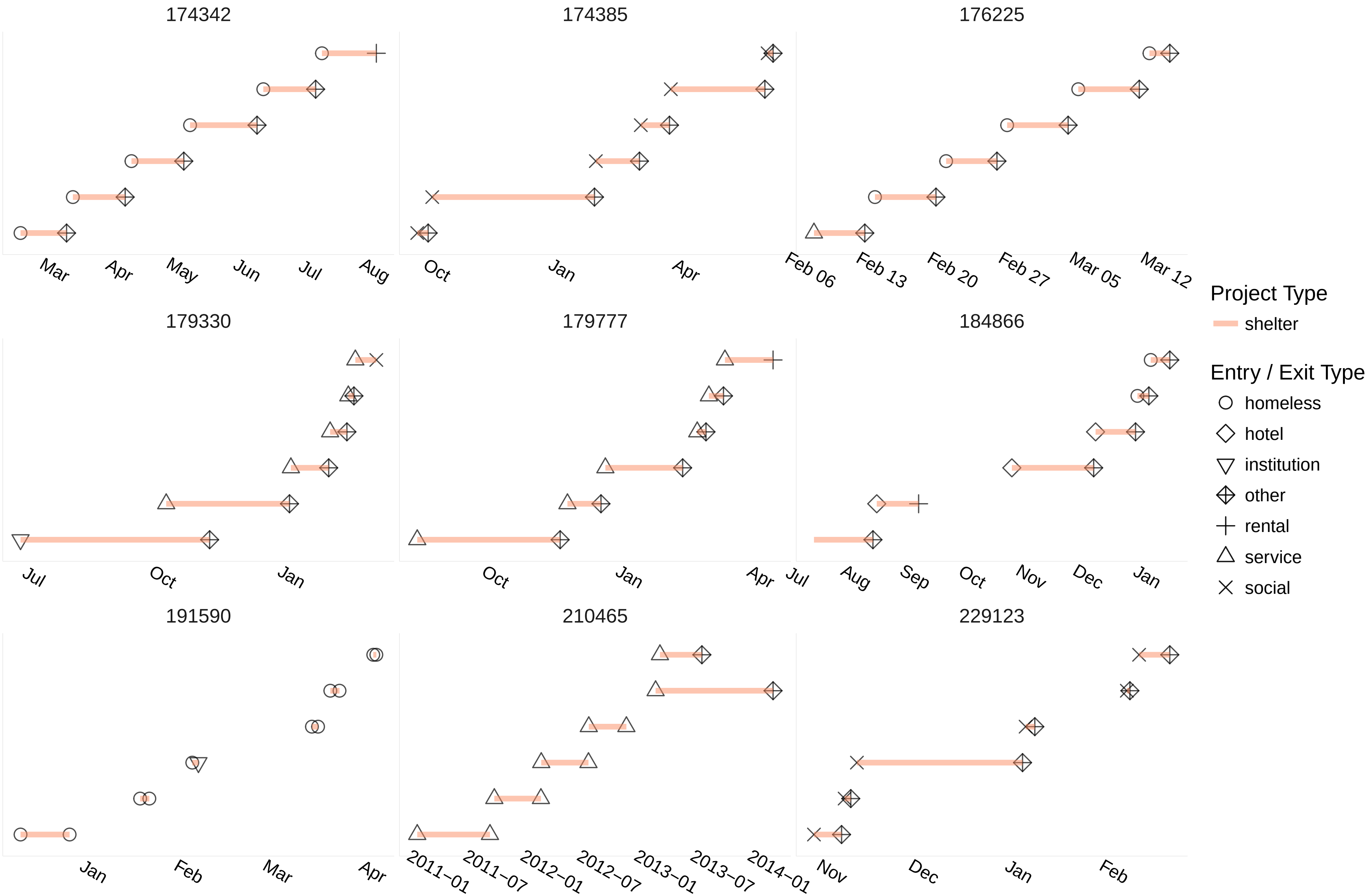

Figure 11: Sample of clients with 6 entry-exit pairs, grouped for consecutive participation in the same project.

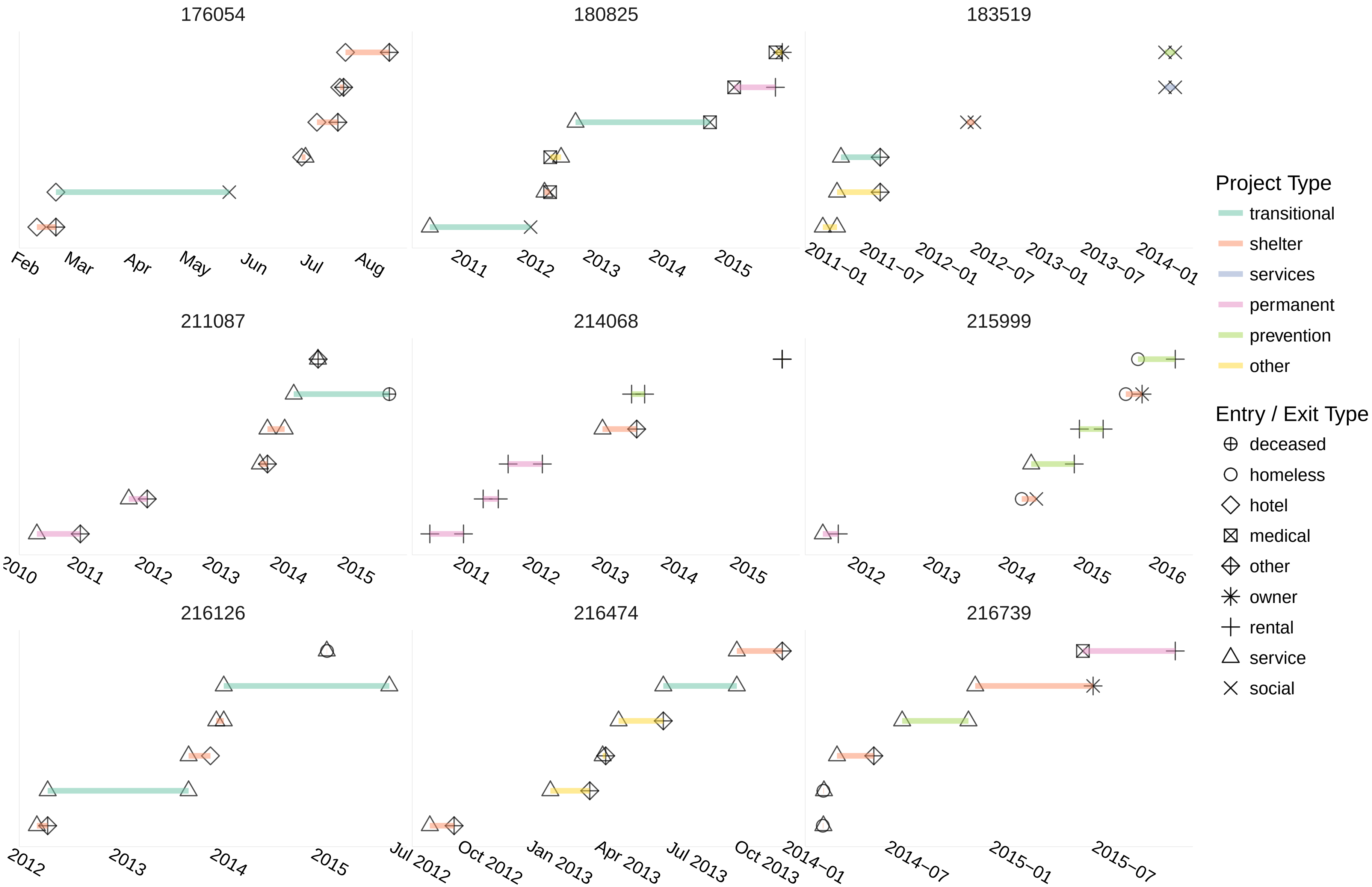

Figure 12: Sample of clients with 6 entry-exit pairs, grouped for consecutive participation in different projects.

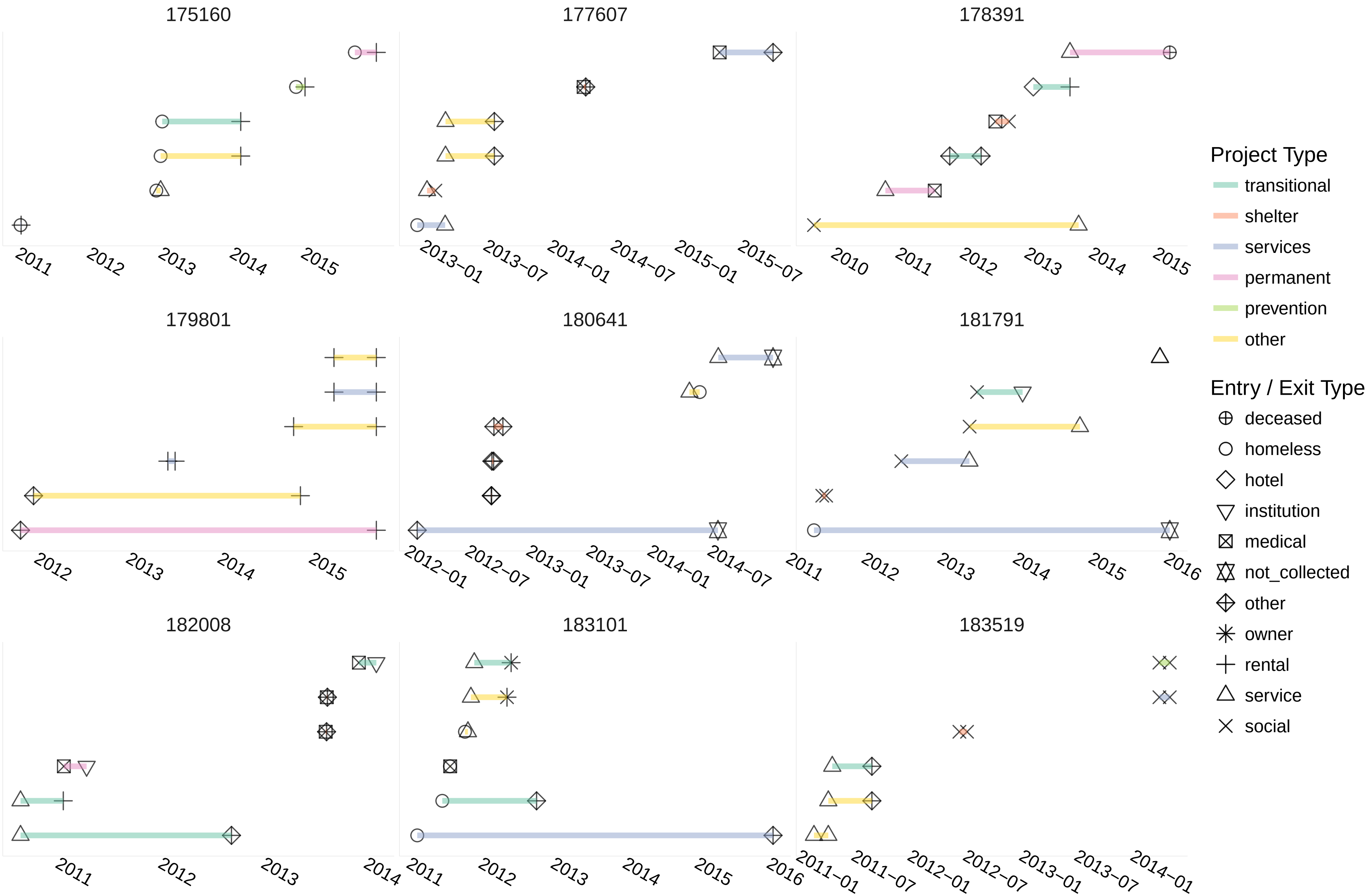

Figure 13: Sample of clients with 6 entry-exit pairs, displaying irregular patterns.

6 Entry-Exit Lengths

The macro view is useful for viewing broad changes over time, but obfuscates variation in the length of the entry-exits, since comparisons in line lengths are difficult when there is so much on the display. As an alternative, we can collapse time and consider just the distribution of entry-exit lengths. With this new focus, we can again study differences across clients and organizations. As in Section 4, we first present the essential “theme” in a single figure, and then develop variations to highlight specific details. However, in this case, the “theme” is the simplest of the plots we show, because it does the most aggregation, and its variations add interesting complexity.

6.1 Overall Entry-Exit Lengths

Figure 14 displays the aggregated entry-exit lengths. To plot is a stacked-histogram of the length of entry-exit pairs across clients, with different exit types colored differently. That is, we could recover 8 different histograms for each of the exit types, if we took out the pieces corresponding to individual colors. The x-axis has bins on the median length of each entry-exit pair, and the y-axis counts the number of pairs that fall into that bin. The bins have been created on a log scale[3].

A point about the statistic we are plotting in this histogram. Remember that each individual can have multiple entry-exit pairs, each with different entry-exit types (e.g., from homeless to service, from service to rental). To account for this, we took the median of the lengths for each entry-exit type used by each person – these are the numbers from which we build the histogram. We used the median because among clients with many entry-exits, they often have many short pairs and maybe one or two long pairs; this approach emphasizes the many short pairs rather than the occasional long pairs. We could have used the maximum, minimum, or average instead, for example, but we had to make a choice, and the median seemed easiest to justify.

Also, we caution that the histograms only include data for entry-exits with endpoints in the categories specified by the legend, so we are only using about 80% of the data set, and inappropriate generalizations should be avoided.

The interpretation of this first entry-exit lengths figure is straightforward. Most pairs that end with exits to service or homeless are finished within a day, this is the large spike at the left. The central blue and purple hill represents entry-exits that end in rental or social after about 100 days. The unexited pairs are most prominent among those that are very long.

[3] Indeed, some entry-exit lengths look like a mixture between a lognormal and a spike at 0.

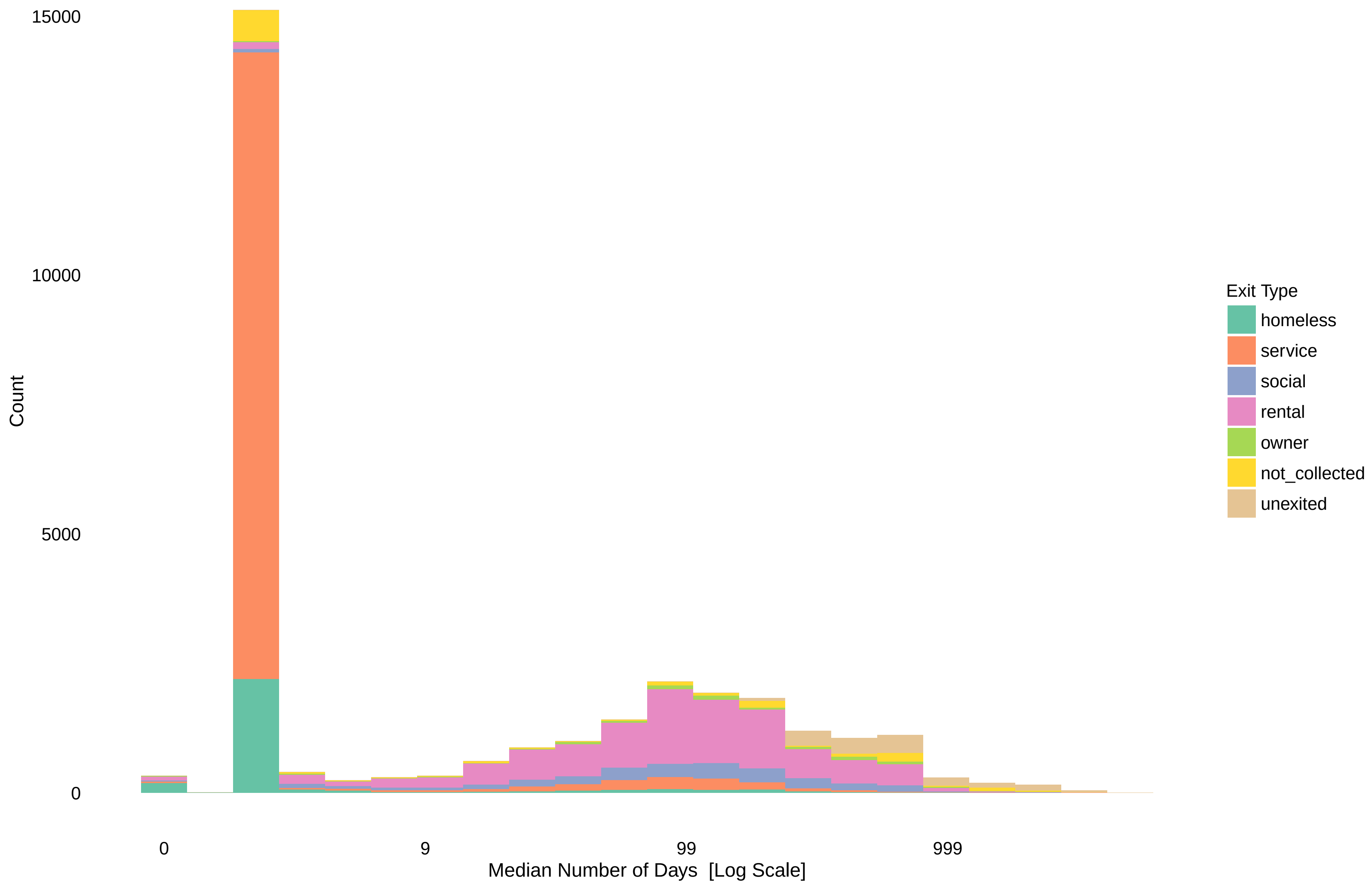

Figure 14: Median length of entry-exit pairs entered after 2010, color-grouped by exit type.

6.2 Entry-Exit Lengths faceted by Entry Type

We can enrich the previous figure by including information on the entry types, which gives us Figure 15. Note that the scale within each cell is not necessarily the same, since some entry types are much more frequent than others.

Some of the patterns suggested in the earlier plot are confirmed in this new view. Among homeless and service entries, most exits take place within a single day; this is true (to a lesser degree) among the social entries as well. Unusually, there seems to also be a spike among those who enter as owners.

We can also compare composition of exits across times. For example, among those who enter as rentals, those who exit within one day tend to exit to service, homeless, or not_collected, while those who have longer entry-exit pairs more frequently exit as rentals. A similar effect can be observed among the other entry types as well. However, given the client interaction lasts longer than a day, the composition of exits seems relatively stable across time lengths. This suggests that entry type and length of stay information would not be very predictive of destination group, after subsetting to only those entry-exits that are longer than one day.

6.3 Entry-Exit Lengths faceted by Project Type

Alternatively, we can split by project type; this is shown in Figure 16.

The other and services categories of programs have fewer records in general, so their histograms appear more heterogeneous. Nonetheless, among the service programs it is possible to make out a large number of rental exits early on, not_collected exits around 100 days, and unexited interactions on the order of several hundred days.

The permanent, prevention, shelter, and transitional housing programs all have relatively well behaved histograms of entry-exit lengths. We notice that most of the entry-exit pairs at shelters last only one day and result in a service exit. Permanent and prevention services both typically last around 100 days; permanent programs tend to lead predominantly to rental exits, while prevention programs lead to a broader mix of exits.

Transitional housing entry-exit pairs are generally longer, and they exhibit a dramatic dropoff in the bin before 999 days, presumably because of the two year limit on transitional housing stays. We note that a larger proportion of the transitional housing exits are actually to social; a larger proportion than among permanent housing programs, for example. Further, we point out that there are more unexited interactions pairs among shorter entry-exit pairs within the transitional cell than in the permanent or prevention cells, which only have unexited interactions when they are long.

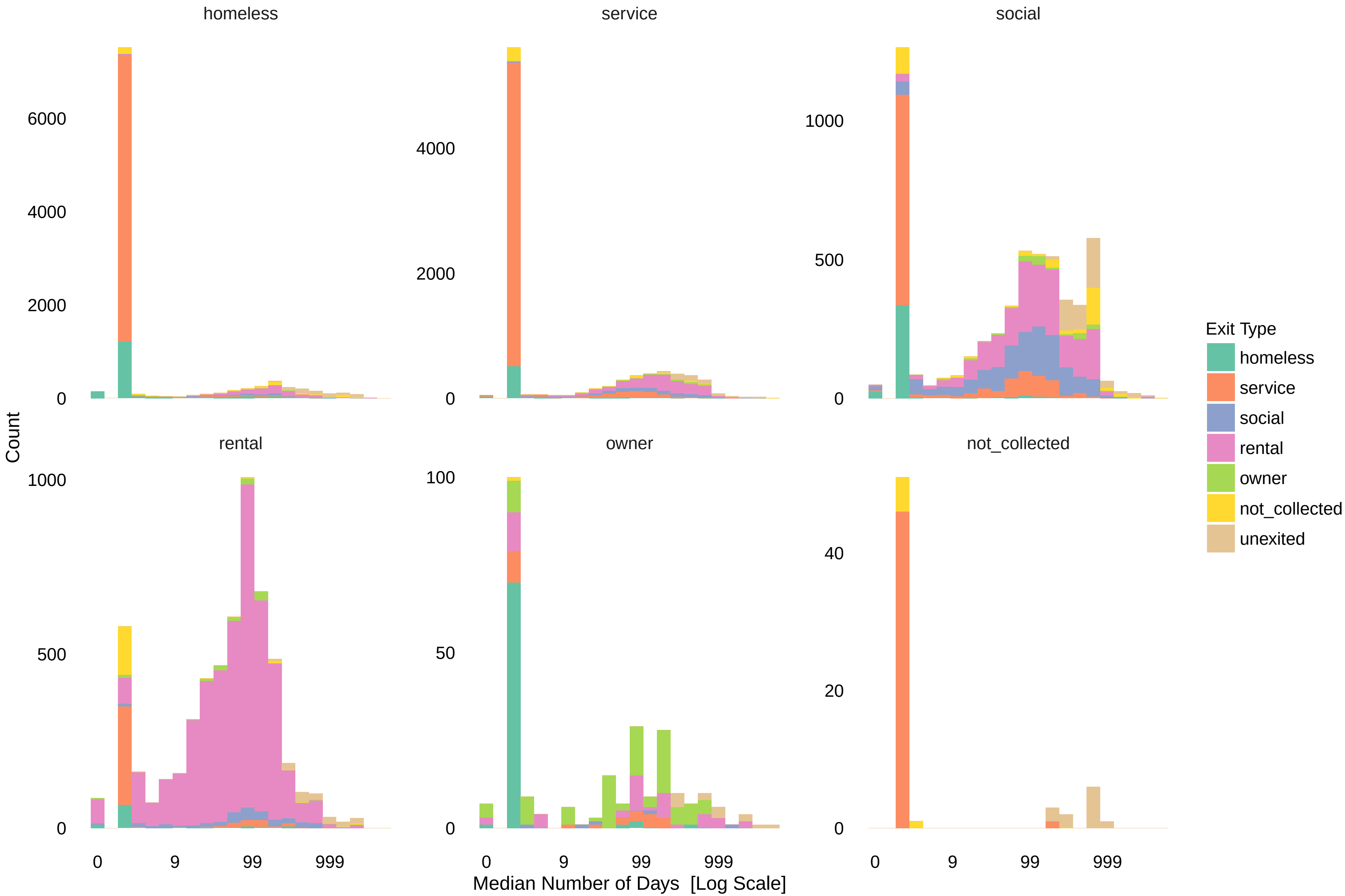

Figure 15: Median length of entry-exit pairs entered after 2010, color-grouped by exit type, faceted across entry types.

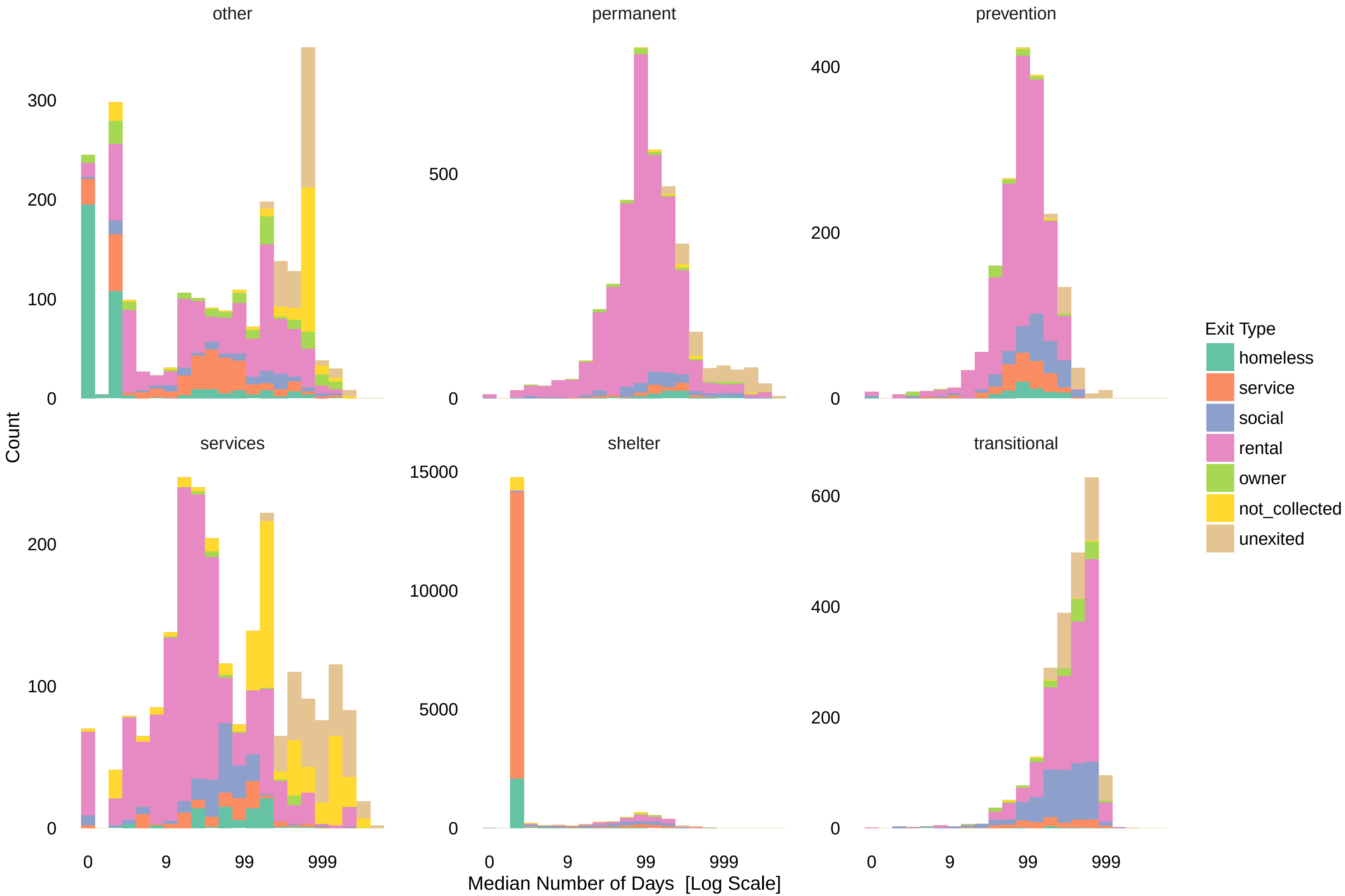

Figure 16: Median length of entry-exit pairs entered after 2010, color-grouped by exit type, faceted across project types.

6.4 Entry-Exit Lengths faceted by Entry and Project Type

We can try to combine the information in the previous two plots, viewing the histogram split by both entry and exit type. The result is in Figure 17. The entry types are listed along the columns, while different projects are along the rows.

There is some interesting variation in entry-exit lengths across project types, given a specific entry type. For example, in the rental entries column, we notice that the services project histogram is centered more to the left, and is also more spread out, compared to the permanent and prevention services.

Studying the "other" projects row, we note an unusual spike in not_collected and unexited exits among those who entered under social. This is most likely related to the mass exits described in the earlier entry-exit plots.

Consider next the homeless entry column. The distribution for "shelter" and "other" organizations are more similar to each other than they are to those for permanent, prevention, and transitional housing projects. But even between shelters and services there are differences – shelters are more likely to have one day exits into services, while "other" projects often have zero day exits back into homelessness.

6.5 Entry-Exit Lengths faceted by Family Membership

To further investigate the differences between families and non-families in the data, we can create histograms faceted by family status. The result is Figure 18. To account for the large differences in the total number of clients who are and aren't in families, we have faceted this figure differently than the others. The reason is that we want the y-axis scales to vary across cells, so that combinations with small counts do not "disappear". Instead of arranging them on a grid which can be read in rows and columns, each cell in this figure must be read independently, and the absolute heights across cells cannot be compared.

Nonetheless, this view makes it easier to compare those who are and are not in families, given a specific entry type. Comparing pairs of horizontally adjacent cells we see the entry-exit lengths of those across family status, given that they have identical entry types. We do caution that some of the cells have very low counts.

Here, we note that families are generally less likely to have short entry-exit pairs, and much less likely to have one or zero day entry-exits. This pattern seems especially true for service and homeless entries. The difference is present, but somewhat less dramatic, among social and rental entries. Indeed, a large proportion rental entries tend to last a long time, even among those not in families.

Further, this view makes it possible to compare the distribution of entry-exit types for families across different entry conditions. For example, we see that those with homeless entries have a higher probability of exiting homeless as well, when compared to service entries – this is the light blue band along the bottom

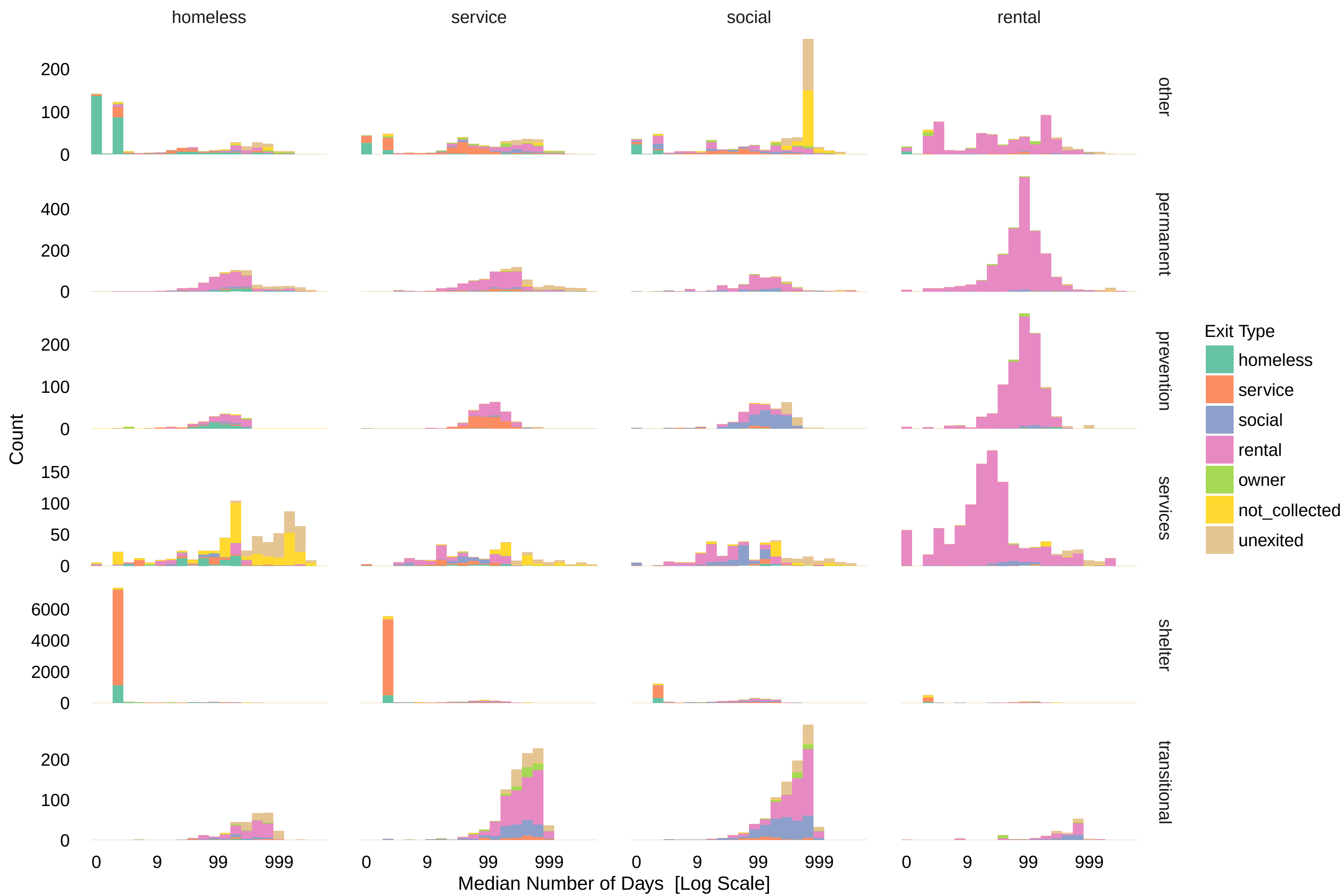

Figure 17: Median length of entry-exit pairs entered after 2010, color-grouped by exit type, faceted across both entry and project types.

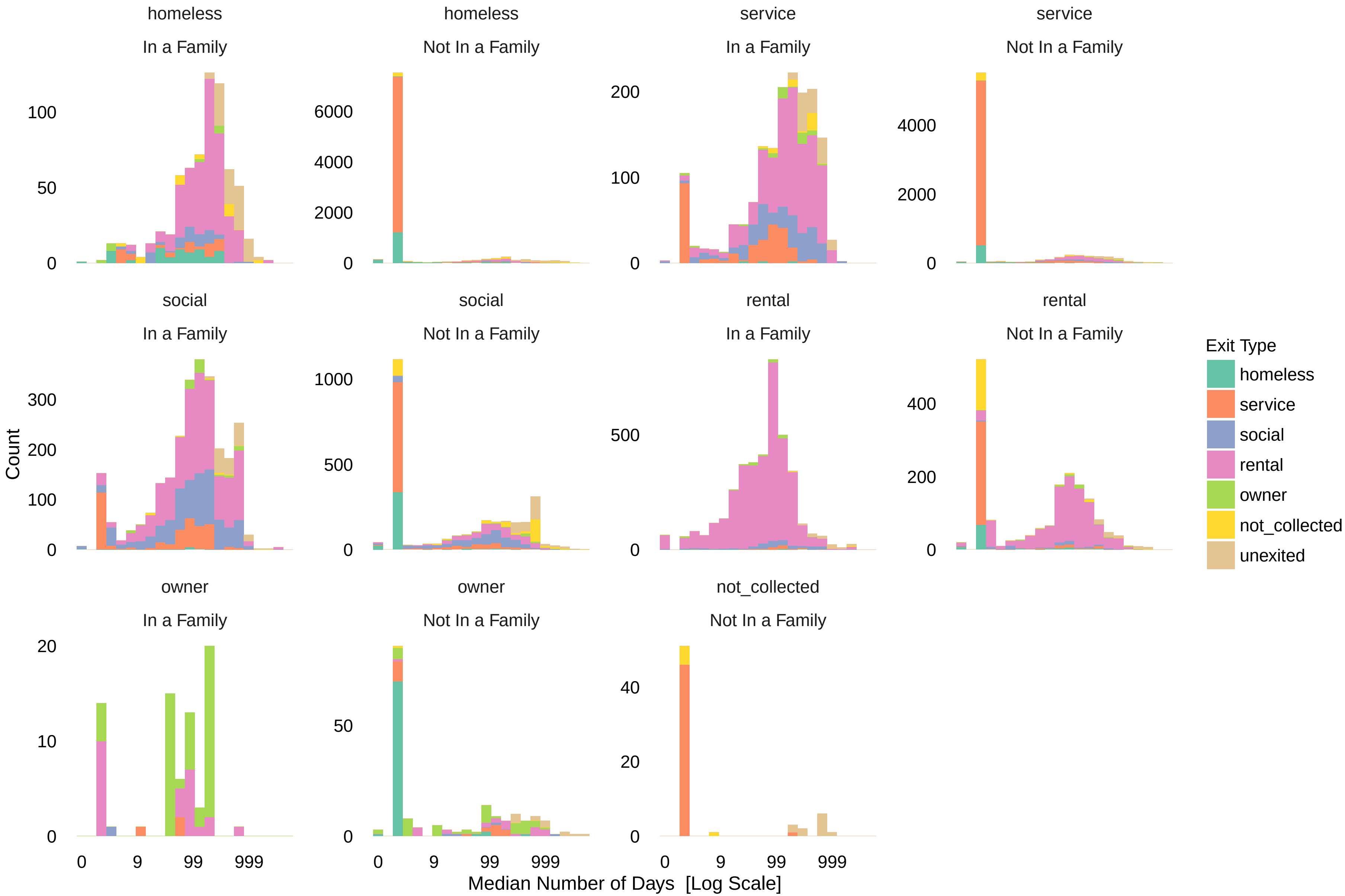

Figure 18: Median length of entry-exit pairs entered after 2010, color-grouped by exit type, faceted by entry type and family status.

of the top left plot. Further, families entering as service and social are more likely to exit in a single day, compared to those entering in homeless or rental types.

6.6 Entry-Exit Lengths faceted by Chronic Homelessness Status

Finally, we thought it would be interesting to facet by the chronic homelessness variable. We also supplement this view with a histogram looking at total time in an entry-exit state, rather than the median as in the previous plots. The results are displayed in Figure 19. The left panel gives the median entry-exit length, while the right gives the total. The top histogram in both panels gives the entry-exit lengths among those who are not chronically homeless, the lower histograms are for those who are.

The main takeaway from the left hand plot is that chronically homeless clients have proportionately more one day entry-exit pairs, and that these one day exits tend to be to homelessness, rather than services. However, considering the total number of days that they are in the system, the overall scale is comparable with the non-chronically homeless (in fact, the mode of the total length of stay among the chronically homeless might be somewhat shorter), but more of their exits are to homelessness. Finally, we note that the 100 day entry-exit spike in the total number of days plot is only visible among the chronically homeless.

6.7 Entry-Exit Lengths shaded by Project Type

For a complementary take on the earlier entry-exit length displays, we consider shading by project types instead of exit types. As in our original entry-exit macro plots, we facet by entry and exit types. Since the general sizes of the counts within each entry-exit cell can vary dramatically, we decided to drop the grid pattern and use a layout similar to the one in the family by entry display. Each cell has two labels, the top describes the entry type, and the bottom gives the exit type.

For example, in the top row, the third cell from the left gives the distribution of lengths of entry-exit pairs among those who enter homeless and exit in rentals. In this cell, the red region are transitional housing projects, as expected these are somewhat longer than the others.

Unsurprisingly, a number of service, not_collected, and homeless exits are for entry-exit pairs that last only one day.

7 Discussion

We hope this document has given illuminating views into the data collected at CTA, and that it can inform future analysis directed towards more specific questions. In particular, we hope that,

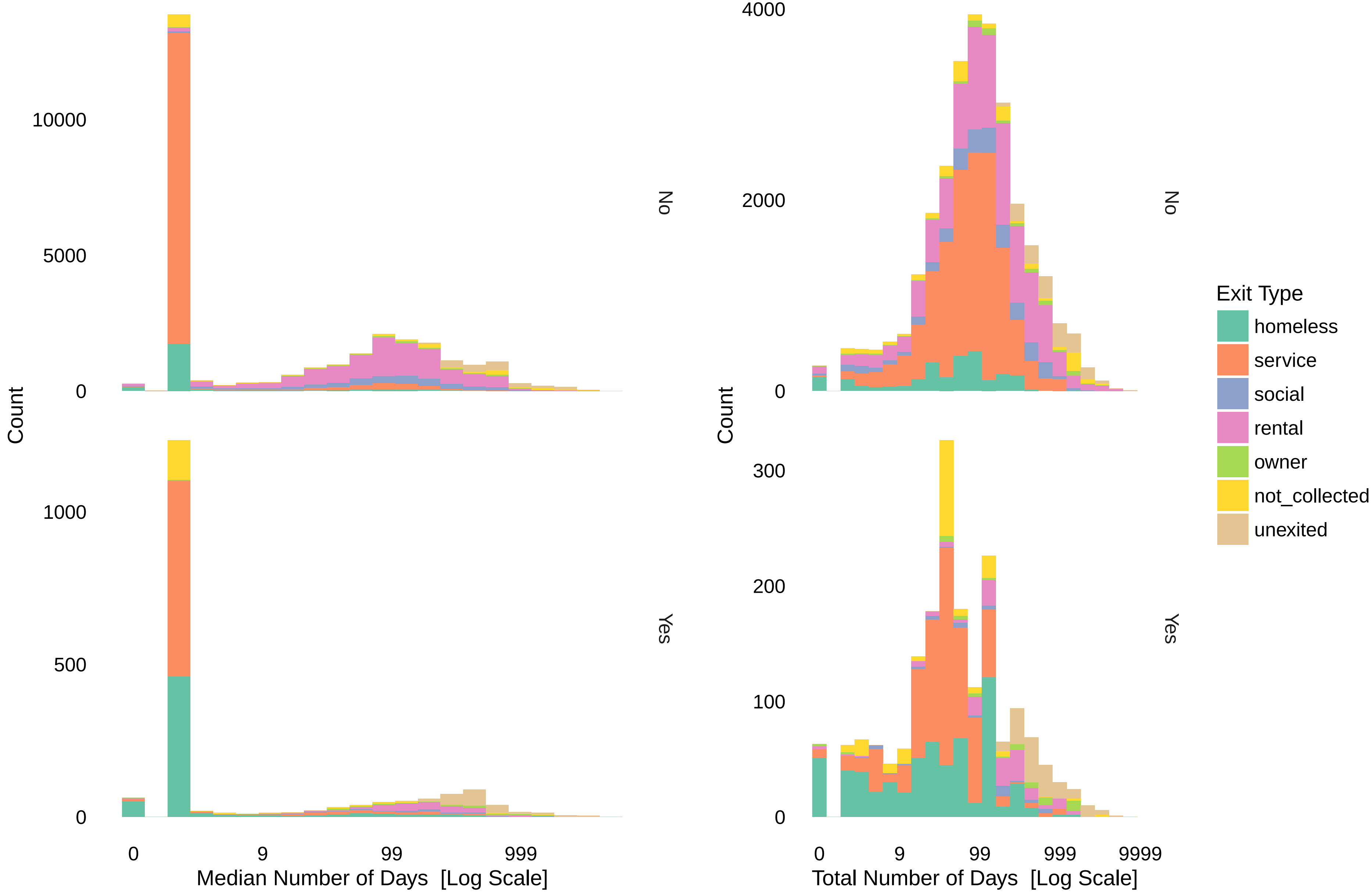

Figure 19: Median (left) and total (right) length of entry-exit pairs entered after 2010, faceted by chronic homelessness status.

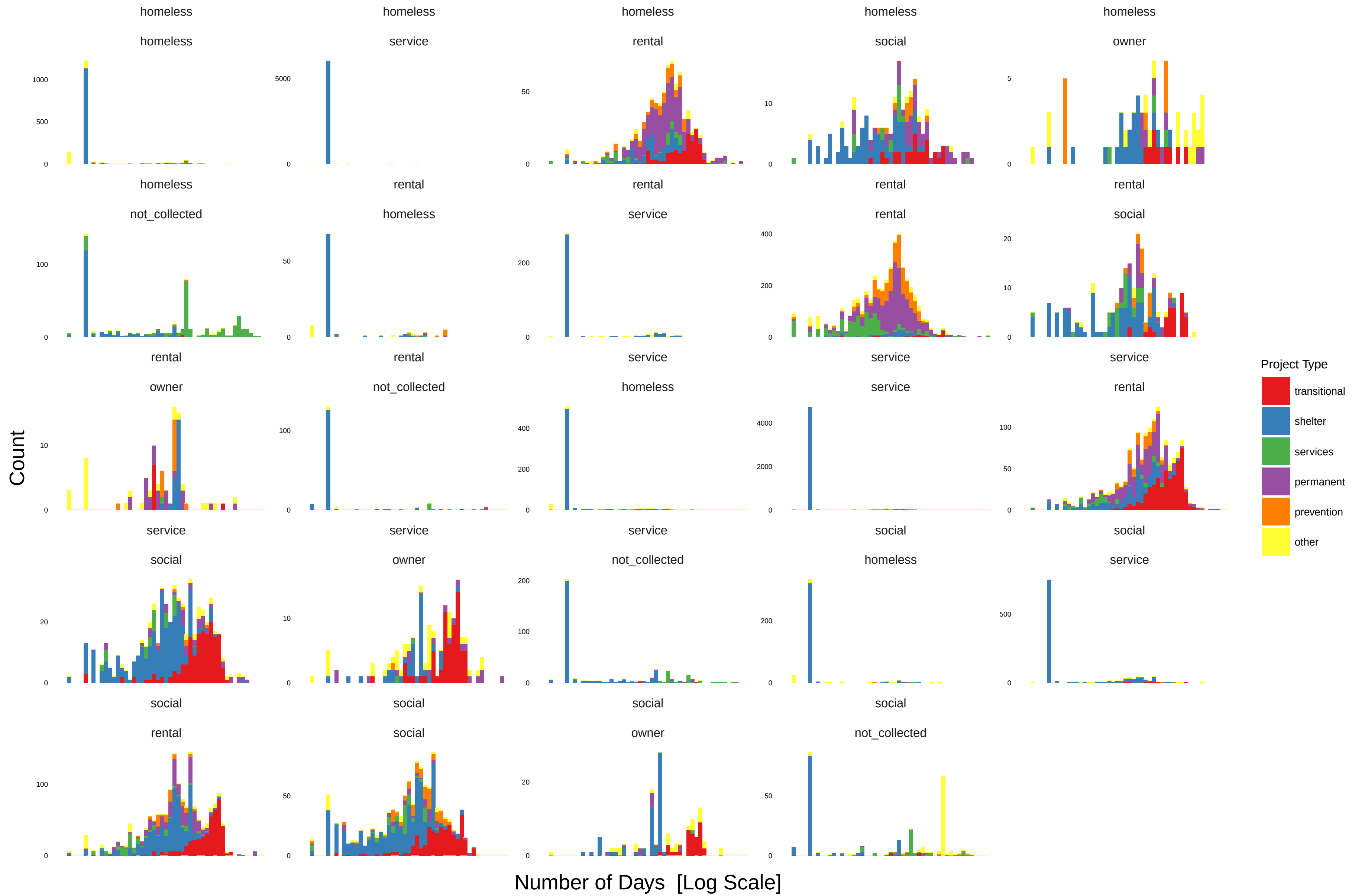

Figure 20: Length of entry-exit pairs entered after 2010, color-grouped by project type, faceted across entry and exit types.

- Analysts have a clearer view of what assumptions may or may not be reasonable, and
- Domain experts might be provoked to generate new, interesting questions based on the available data.

Again, all figures in this document are reproducible from a single script. Other brainstorming and preprocessing code is available on our private code repository, if you would like access, don't hesitate to ask. Finally, we welcome any follow-up and ideas to direct future analysis.

8 Appendix

8.1 Data Anomalies

While preprocessing the original data, we came across several unexpected details. These are most likely just artefacts in data collection, and the probably have no repercussions for answering most of the questions of interest. Nonetheless, we thought we would record them here, so that they do not surprise future analysts.

- The first surprise we encountered was that the "Housing Status @ Project Start" and "Living situation before program entry?" fields were not entirely consistent with one another. This can be seen from Figure 21, which plots the number of responses to "Living situation" after subsetting to responses to "Housing Status." For example, in the second row (Housing Status in "Category 1," which is homelessness), we see that, while most of the "Living situation" responses are set to "emergency shelter" or "place not meant for habitation," a number lie in other categories that we would not usually associate with homelessness. Similarly, for the final row "Stably Housed," we notice a number who are living in emergency shelters, according to the "Living situation" question.
- In the entry-exit table, the entry income and exit income features are essentially equal. This is evident in Figure 22. They do not seem completely identical – there are some differences of less than a dollar – which makes us think it might not be a purely data handling error. On the other hand, we are suspicious of whether the exit variable can be trusted. Does the system default to setting exit income equal to entry income in most cases?
- The relationship between veteran ages and year of entry into the military also seems suspect, as displayed in Figure 23. In particular, the points boxed in red seem unlikely to be correct. For the top rectangle, it seems unlikely that older clients have just joined the military, while for the middle left square, the client would not have been born in time to enter in the specified year.

- This is not so much an anomaly as point we were not informed about earlier, but we noticed that when the raw data files say that the data are from “after 2010,” it means that the exits were recorded after 2010. There are in fact a number of entries before 2010, as displayed in Figure 24.
- Again, this is not necessarily an anomaly, but the project_id column is not unique across the raw projects table – it sometimes duplicated when the same organization serves multiple cities.

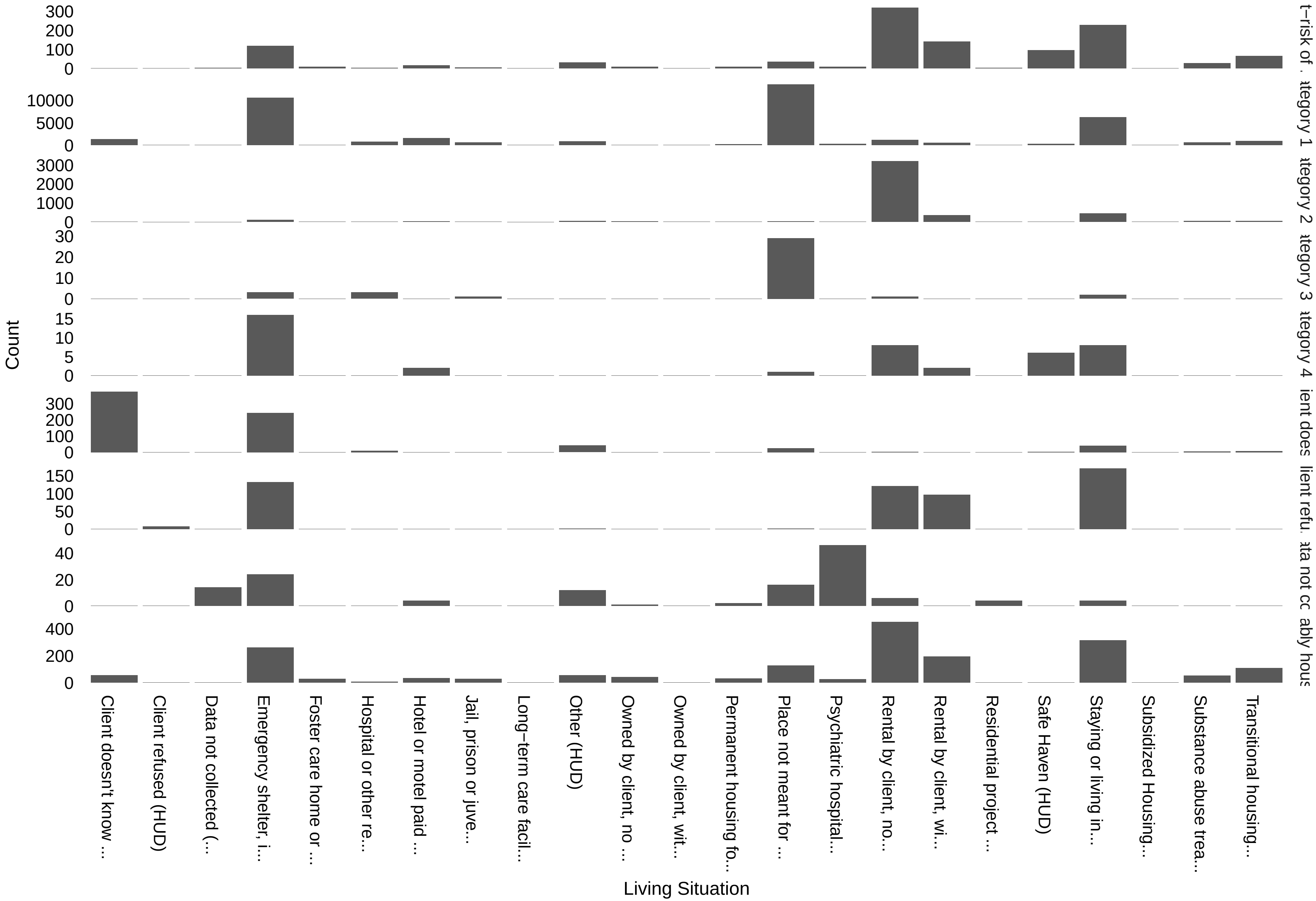

Figure 21: The "Housing Status @ Project Start" and "Living situation before program entry?" fields are not entirely consistent.

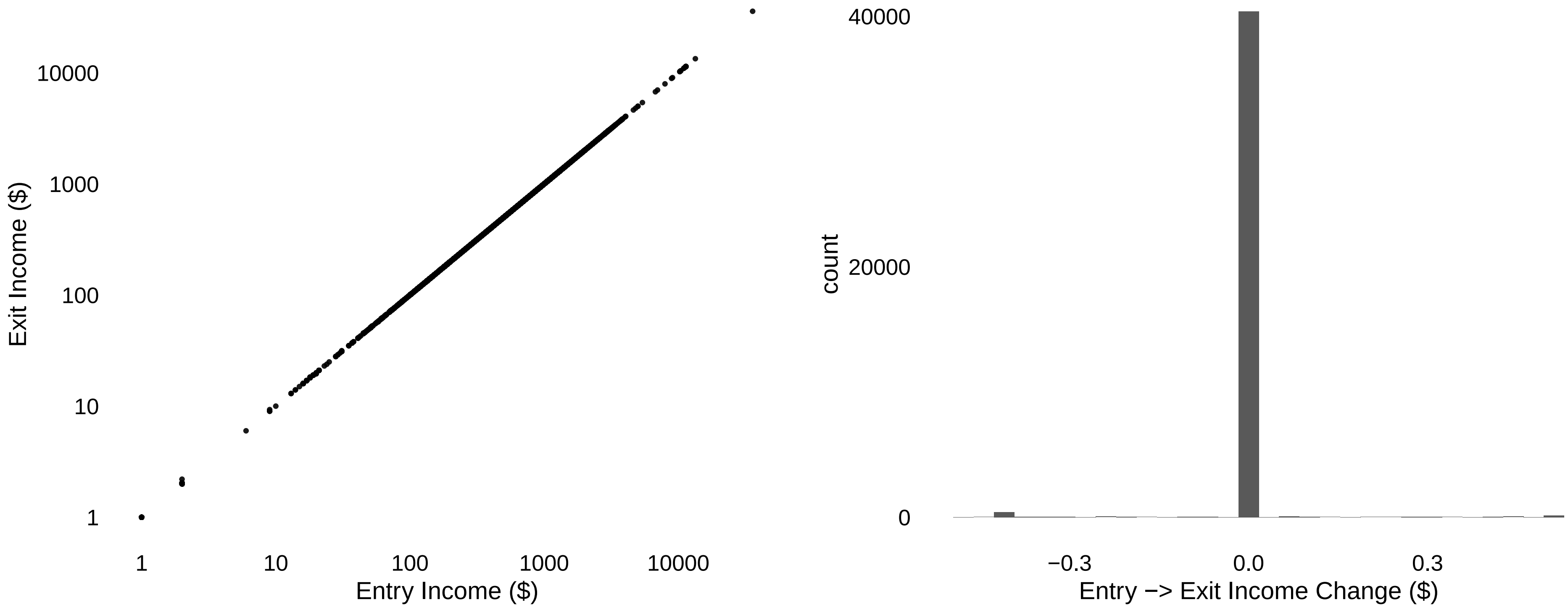

Figure 22: The left hand side plots entry income against exit income, on a log scale, whenever both variables are available. The right hand gives a histogram of the difference between entry and exit income.

Figure 23: The relationship between client age and veteran year of entry seems suspect, at least for the points annotated by red rectangles.

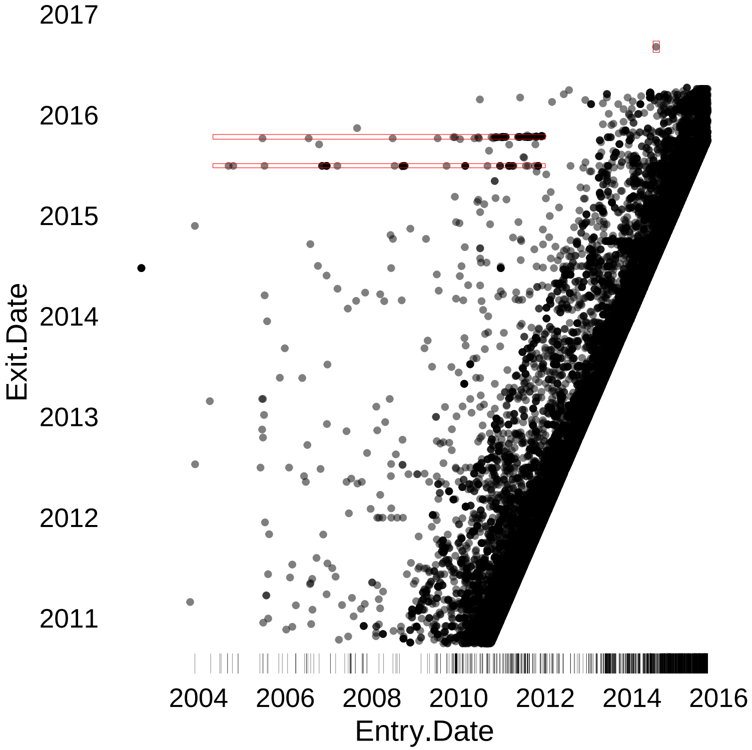

Figure 24: The data are clients with *exits* after 2010. Also, one person has an exit in the future, though this is not impossible – in theory the service has an exit date planned. Also, we have highlighted two bands that seem to correspond to mass-exits within small time frames. The small lines along the bottom represent entry dates with no recorded exit dates.