

Fellows: Io Flament, Cristina Lozano, Momin M. Malik¹ Project Manager: Laura Szczuczak
Technical Mentors: Qiwei Han², Euro Beinat³, Dario Bertocchi⁴, Bartosz Hawelka³, Pavlos Kazakopoulos³

¹Carnegie Mellon University, ²Nova School of Business and Economics, ³University of Salzburg, ⁴Università Ca' Foscari Venezia

The Problem

Changing patterns of tourism have put unprecedented pressure on the infrastructures and social makeup of cultural heritage sites around the world.

Florence, one of the world's most famous destinations, fills with tourists that crowd the city center, changing economic patterns and making life difficult for city residents and for the tourists themselves. On an average day, tourists account for about $\frac{1}{3}$ of the people seen in the city (fig. 1).

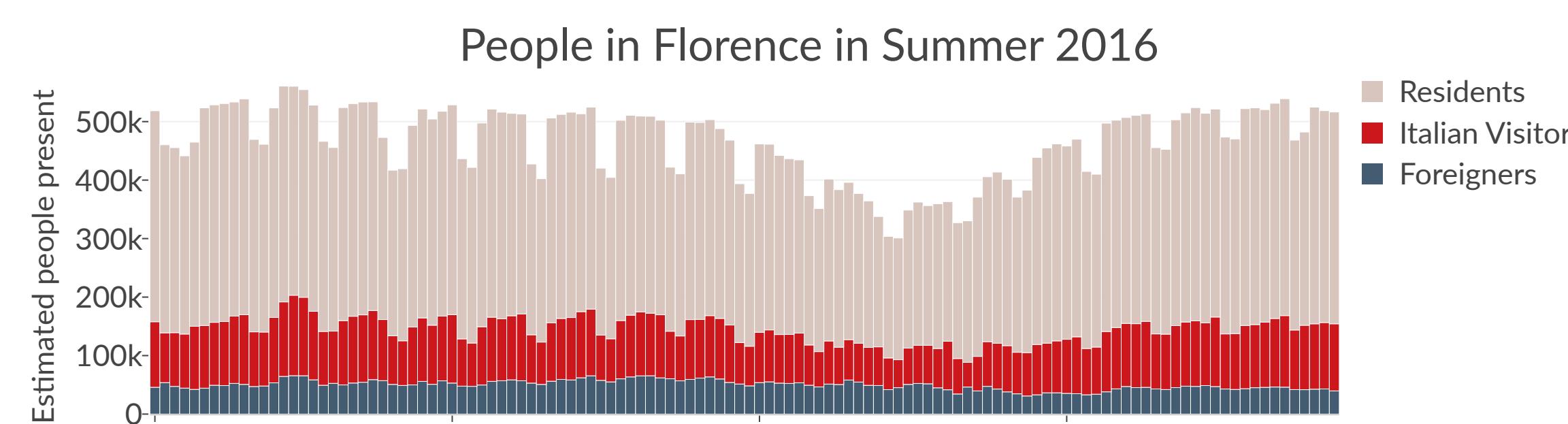


Figure 1. Daily number of people present in the city of Florence, estimated from telecom data.

The future of tourism in Florence looks unsustainable for the city. Meanwhile, other cities and towns across Tuscany, and even smaller museums in Florence, are underutilized. And the lack of detailed data about tourists movements makes it hard to make policy-related decisions.

Our Approach

Florence is awash with constant digital data, like traces of calls passing through cell phone towers and logs of museum entrances. We take a novel approach of applying this data to study tourism, greatly enriching what we see in standard administrative data like overnight stays and airport arrivals.

We link these underutilized sources of data to better understand patterns of travel, stay, and activity in the city of Florence and region of Tuscany. With a more granular picture of what is happening in the city, the government can better serve tourist populations with infrastructure and information while preserving the quality of life of residents.

The Data

Our project uses the following data for the months of June through September, 2016:

300M+ anonymized Call Detail Records (CDR)

390K Firenze card (Florence's museum pass) entry logs

Aggregate total monthly museum entries

Airport arrivals

Tourist center visits

Hotel presences and locations

Alone, each source gives a partial picture of tourist activity in Florence. But by triangulating findings from different sources of data, we can both get a better picture of tourism and inform future systems of automatic monitoring and dynamic management.

Methodology

To look at crowding, we:

- Create dynamic network edge lists for *transitions* (consecutive visits) between both museums from Firenze card logs and cell towers coverage areas from Call Detail Records (CDR)
- Aggregate to create first-order Markov chains
- Create network visualization of the graph of transitions
- Use the transition graph to make interactive visualization

For finding patterns in Firenze card usage:

- In logs, we see many consecutive Firenze cards ID numbers visiting the same museums in the same order, strong evidence for group travel
- To identify groups, we calculate Levenshtein edit distance between ordered lists of museums visited by pairs of Firenze cards (small edit distances compared to the total path length)

Findings

We focus on three findings: our measurements of which museums account for which share of total movements between museums, a breakdown of group behavior in card usage, and patterns that reveal motivations for buying the Firenze cards.

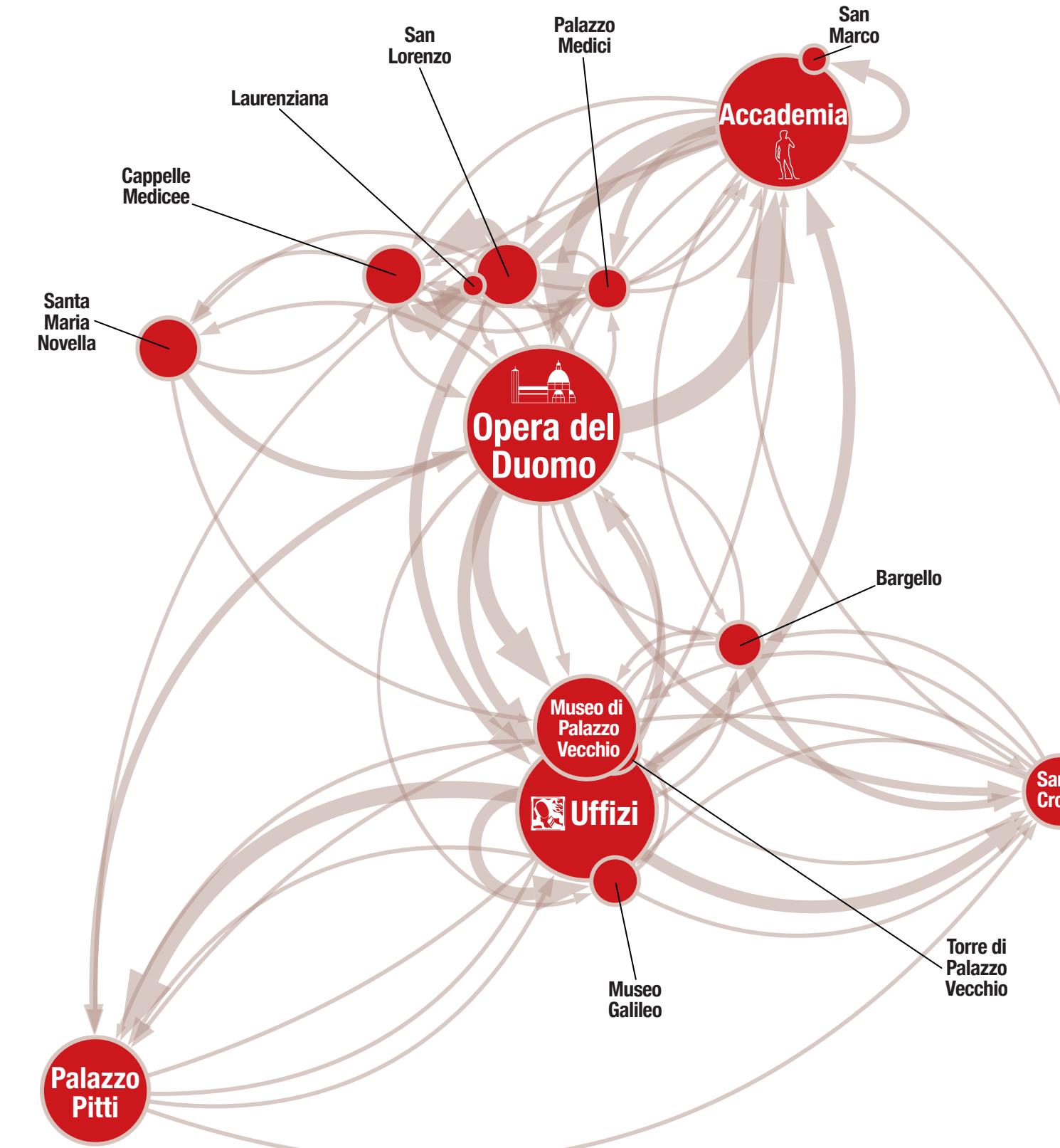


Figure 2. The network of Firenze card sites, and flows between them. Nodes are scaled to total number of visitors, and edges are scaled to the number of people going from one site to another on a given day using the Firenze card.

Size of group	Children with group			
	0	1	2	3+
1	22%	2%	<1%	<1%
2	53%	3%	2%	<1%
3+	15%	<1%	<1%	<1%

Table 1. Breakdown of total Firenze card users, Summer 2016, by group size and number of children. The main type of Firenze card usage is two adults, traveling together without children. This also shows that people in large groups account for a tiny fraction of overall card users.



Figure 3. The most frequent Firenze card usage is to visit several of the overall most popular museums. Surprisingly, many cardholders see museums all on one day rather than spread over the 72 hours of the Firenze card, which gives an idea of how many people buy the Firenze mainly to skip lines (and not for economy).

Project outputs

- Written report on tourist movements in Florence
- Recommendations for data collection practices, sharing, and integration
- Public, open-source code repository and documentation
- Interactive visualizations of patterns and flows:

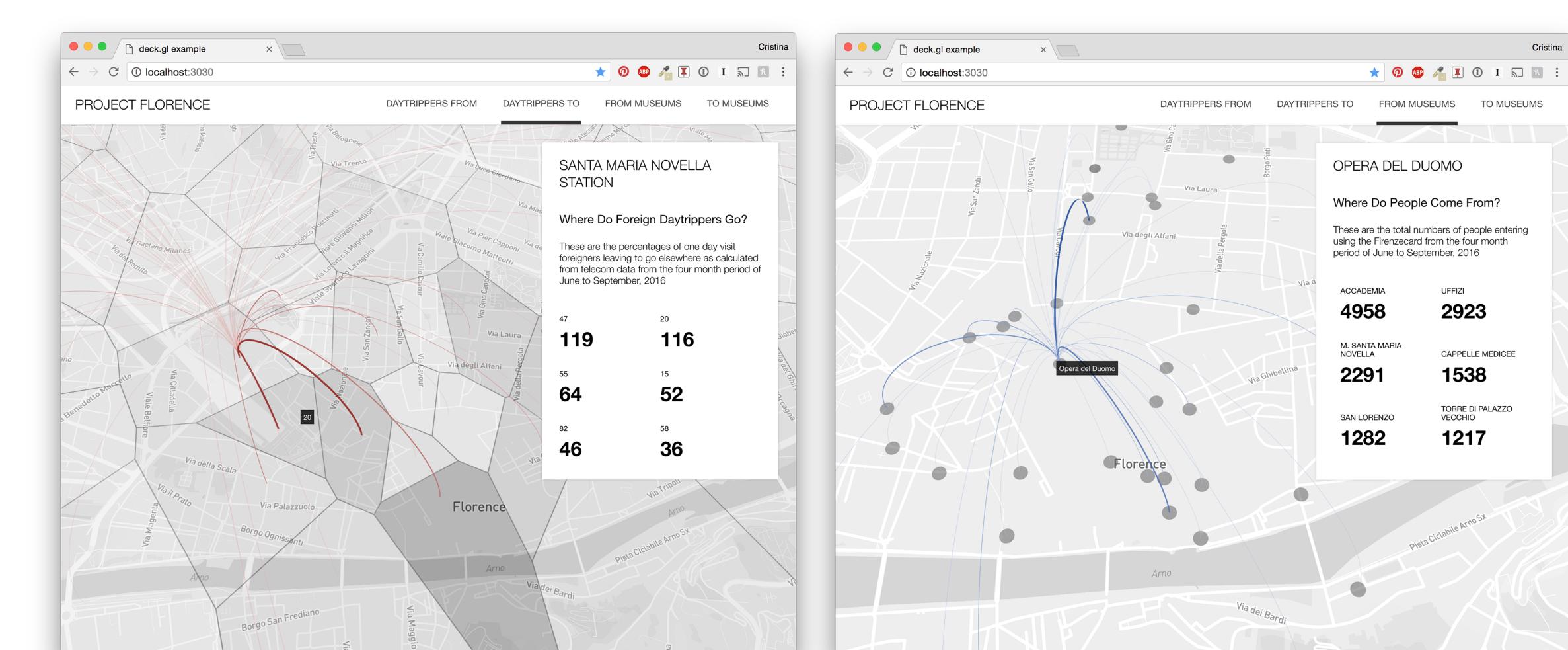


Fig 4. Screenshots of interactive visualizations for pairwise flows between cell phone towers (left) and museums (right).

Recommendations

- Predict crowding and give tourists dynamic counters, available online, to enable people to avoid queues and thereby reduce crowding
- Make recommendations to tourists (in-person or via app) about alternative sites to visit, or alternative orders in which to visit the same sites in ways to distribute crowds
- Continue building up data collection, sharing, and analysis procedures between agencies

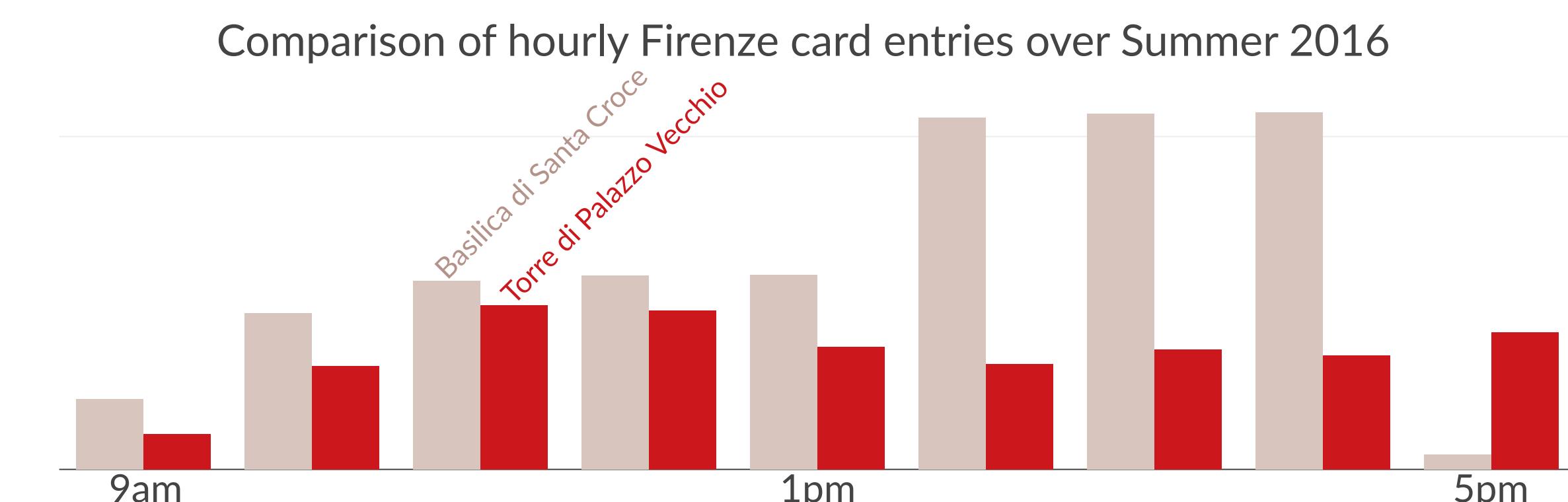


Figure 5. This comparison suggests a large difference in the sizes of crowds aggregating at museums in the afternoon. Such findings can be used to make recommendations to reduce waiting lines and crowding of streets and public spaces.

With these types of measures, Tuscany and the city of Florence can meet the challenges of changing patterns of tourism head-on, accommodating increasing numbers of visitors while preserving both the quality of tourist services and the quality of life for locals. It can serve as a model for other locations worldwide who have begun to similarly experience unsustainable tourism.

Partners and data providers

