Table 2: Map statistics for maps currently supported by `trajdata`. Note that the INTERACTION dataset does not provide pedestrian walkway or crosswalk information [39].

Dataset	Lane Length (km)	Road Area (m^2)	Pedestrian Area (m^2)
nuScenes [18]	212.85	946,275	250,164
INTERACTION [39]	18.78	76,502	–
Lyft Level 5 [19]	185.42	591,333	17,359
nuPlan [27]	325.95	1,327,965	271,277

Table 3: Proportion of stationary agents per dataset. S/M denotes Single/Multi.

Dataset	Proportion	Dataset	Proportion
ETH [22]	4.0%	INTERACTION S/M [39]	5.2%/4.5%
UCY [23]	0.0%	Lyft Level 5 [19]	0.1%
SDD [40]	5.1%	Waymo Open [17]	53.6%
nuScenes [18]	17.5%	nuPlan [27]	0.1%

A Additional Dataset Details

Licenses. The ETH [22] and UCY [23] datasets are provided for research purposes[4]. SDD [40] is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 (CC BY-NC-SA 3.0) License. nuScenes [18] and nuPlan [27] are mostly licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License (CC BY-NC-SA 4.0), with modifications outlined in `https://www.nuscenes.org/terms-of-use`. The INTERAC-TION [39] dataset is provided for non-commercial teaching and research use[5]. The Lyft Level 5 [19] dataset is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 license (CC-BY-NC-SA-4.0).The Waymo Open Motion Dataset [17] is licensed under its own Waymo Dataset License Agreement for Non-Commercial Use[6].

Protecting Personal Privacy. All datasets supported by `trajdata` are captured in public spaces. Further, each of the ETH [22], UCY [23], SDD [40], and INTERACTION [39] datasets capture data in public spaces from elevated fixed traffic cameras or drones, whose birds-eye viewpoints shield the faces of pedestrians and drivers from being collected. The nuScenes [18], Lyft Level 5 [19], Waymo Open Motion [17], and nuPlan [27] datasets each preserve privacy by leveraging state-of-the-art object detection techniques to detect license plates and faces.

B Map Statistics

As shown in Table 2, current AV dataset maps are very large, spanning multiple neighborhoods in different cities. The INTERACTION [39] dataset is a notable exception in magnitude, however, due to a drone camera's limited spatial observation range.

C Stationary Agents

Table 3 summarizes the amount of stationary agents per dataset. As can be seen, nuScenes (17.5%) and Waymo Open (53.6%) are comprised of many parked vehicles.

---

[4]See the statement at the top of `https://icu.ee.ethz.ch/research/datsets.html` and in the "Crowds Data" card of `https://graphics.cs.ucy.ac.cy/portfolio`.
[5]See `http://interaction-dataset.com/terms-for-non-commercial-use` for full terms.
[6]Full terms can be found at `https://waymo.com/open/terms/`.

14
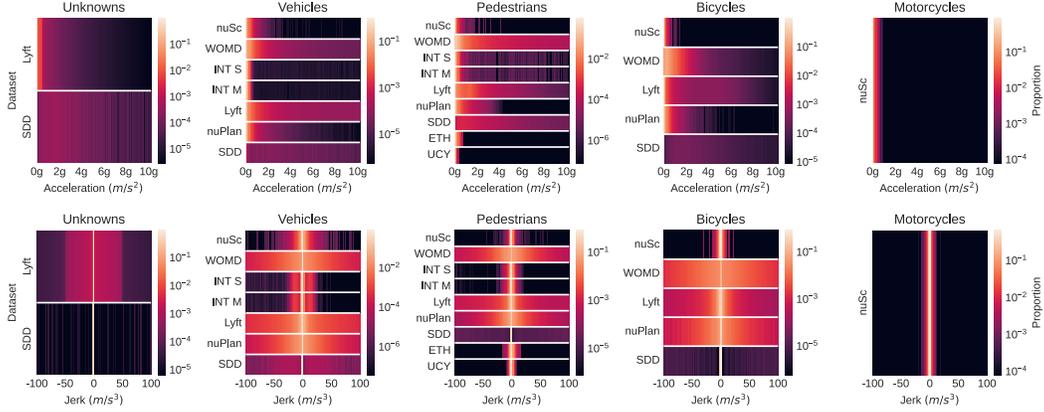
Figure 12: Acceleration (**top**) and jerk (**bottom**) distributions.
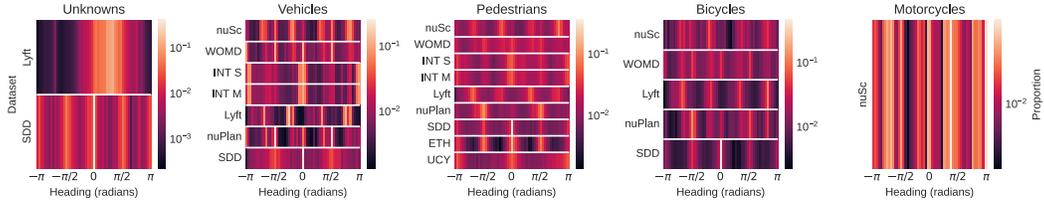
Figure 13: Unnormalized agent heading distributions per dataset and agent type.

D   Acceleration and Jerk

Similar to the speed distributions in the main text, Fig. 12 shows that Waymo [17], Lyft [19], and
SDD [40] have long-tailed acceleration and jerk distributions. Further, as seen in the main text, the
fixed-camera-based ETH [22] and UCY [23] datasets, as well as the drone-based INTERACTION
dataset [39], yield tightly-clustered distributions around smooth motion (generally having small
acceleration and jerk magnitudes). Note that these values are derived by `trajdata` via finite
differencing. Accordingly, some overestimation of the acceleration and jerk distribution supports are
to be expected.

E   Heading Distributions

The distributions of (unnormalized) agent headings are shown in Fig. 13. As can be seen, most
distributions contain peaks around 0 and $\pm\pi/2$ radians, as north-south and east-west roads are very
common in many cities. As a particular example, Fig. 14 visualizes heading distributions for vehicles
and pedestrians in the Waymo Open Motion Dataset [17], showing that pedestrians have much more
varied heading values than road-following vehicles.

F   Path Efficiency

As can be seen in Fig. 15, most path efficiency distributions are uniformly distributed, with peaks
near 100% (shown as brightly-colored regions), echoing earlier straight-line findings. Further, the
INTERACTION [39] dataset is an outlier in that its vehicle and pedestrian trajectories are virtually
all straight lines with much less curved motion than other AV and pedestrian datasets.

15
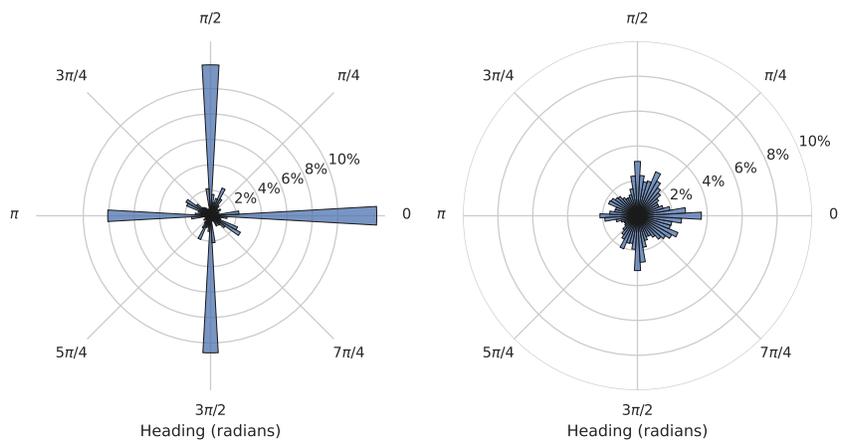
Figure 14: Vehicle (**left**) and pedestrian (**right**) heading distributions in Waymo Open [17].
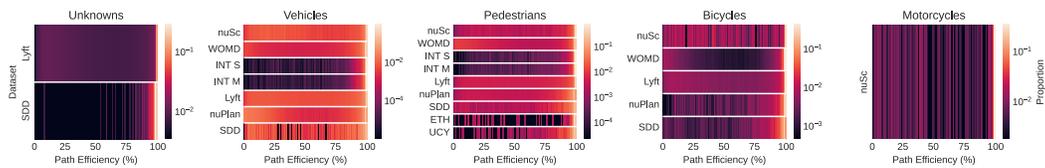
Figure 15: Path efficiency distributions visualized per agent type and dataset.

16