Customer Lifetime Value: Models, Metrics and a Multitude of Uses

PyData Los Angeles, Oct 2018

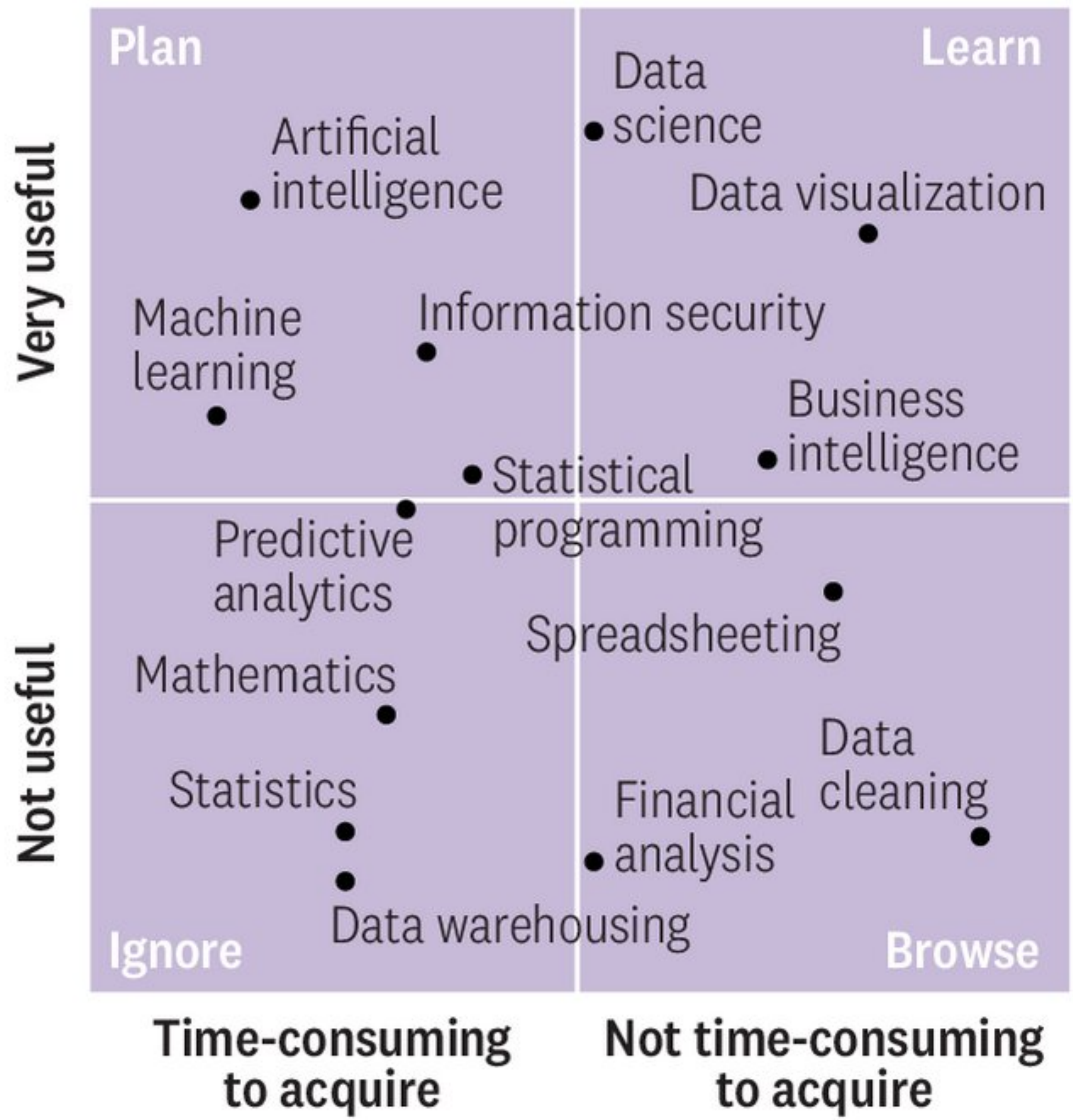
Brian Bloniarz

SECOND MEASURE

Outline

1. Analyzing companies at the grain of a customer
2. Initial look at customer lifetime spend [metrics]
3. Estimation [models]
4. Models, applied to data [uses]
5. Questions?

An Example of How to Plot Data Skills on a 2x2 Learning Matrix



Source: Analysis of internal data learning needs by Filtered

HBR

I. Framework

What happens when you set the unit of analysis to be a customer?

You start to think about all the accounting cash flows associated with that customer...

- Purchases (Revenue)
- Cost of goods sold (COGS)

What happens when you set the unit of analysis to be a customer?

You start to think about all the accounting cash flows associated with that customer...

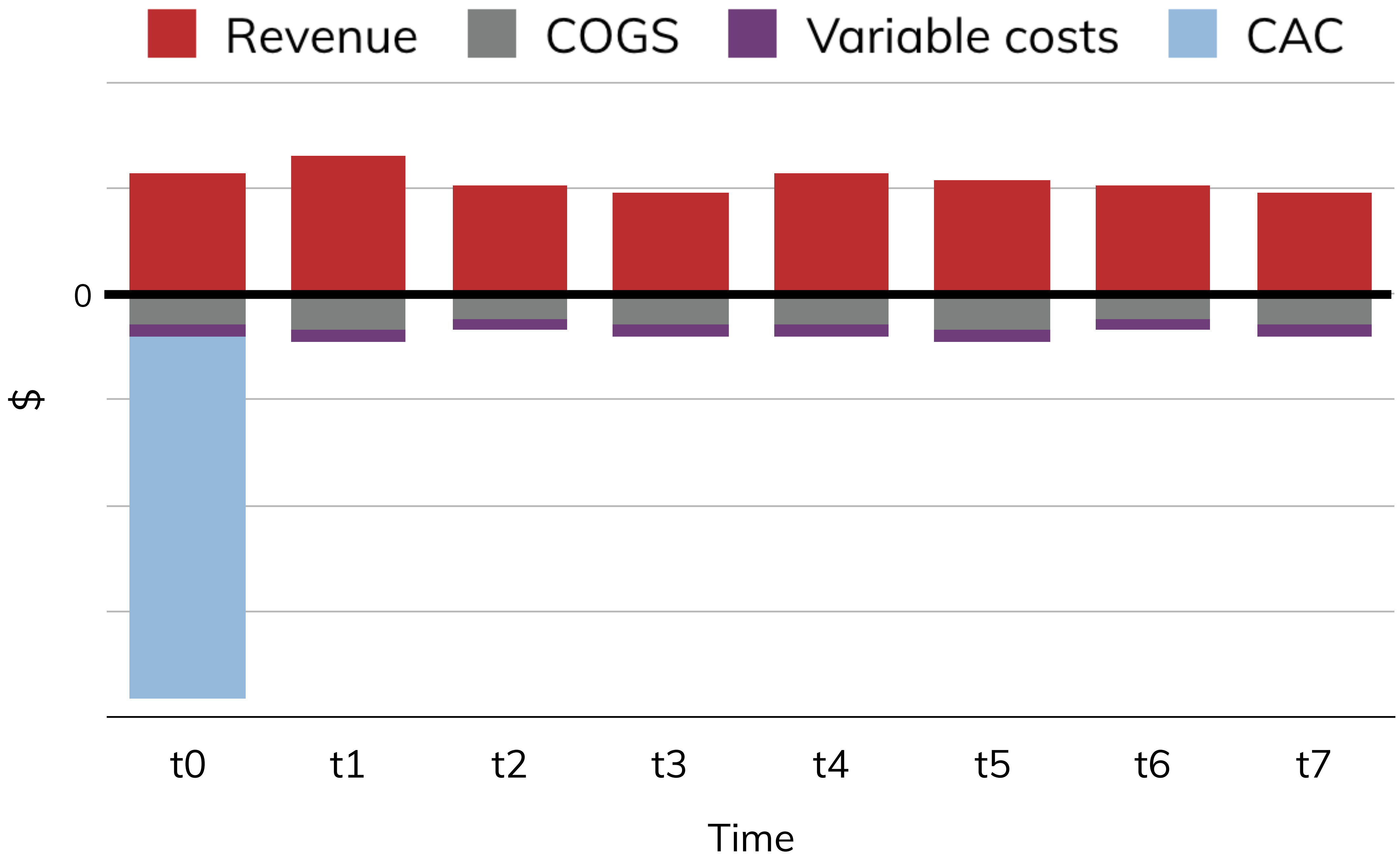
- Purchases (Revenue)
- Cost of goods sold (COGS)
- Variable costs

What happens when you set the unit of analysis to be a customer?

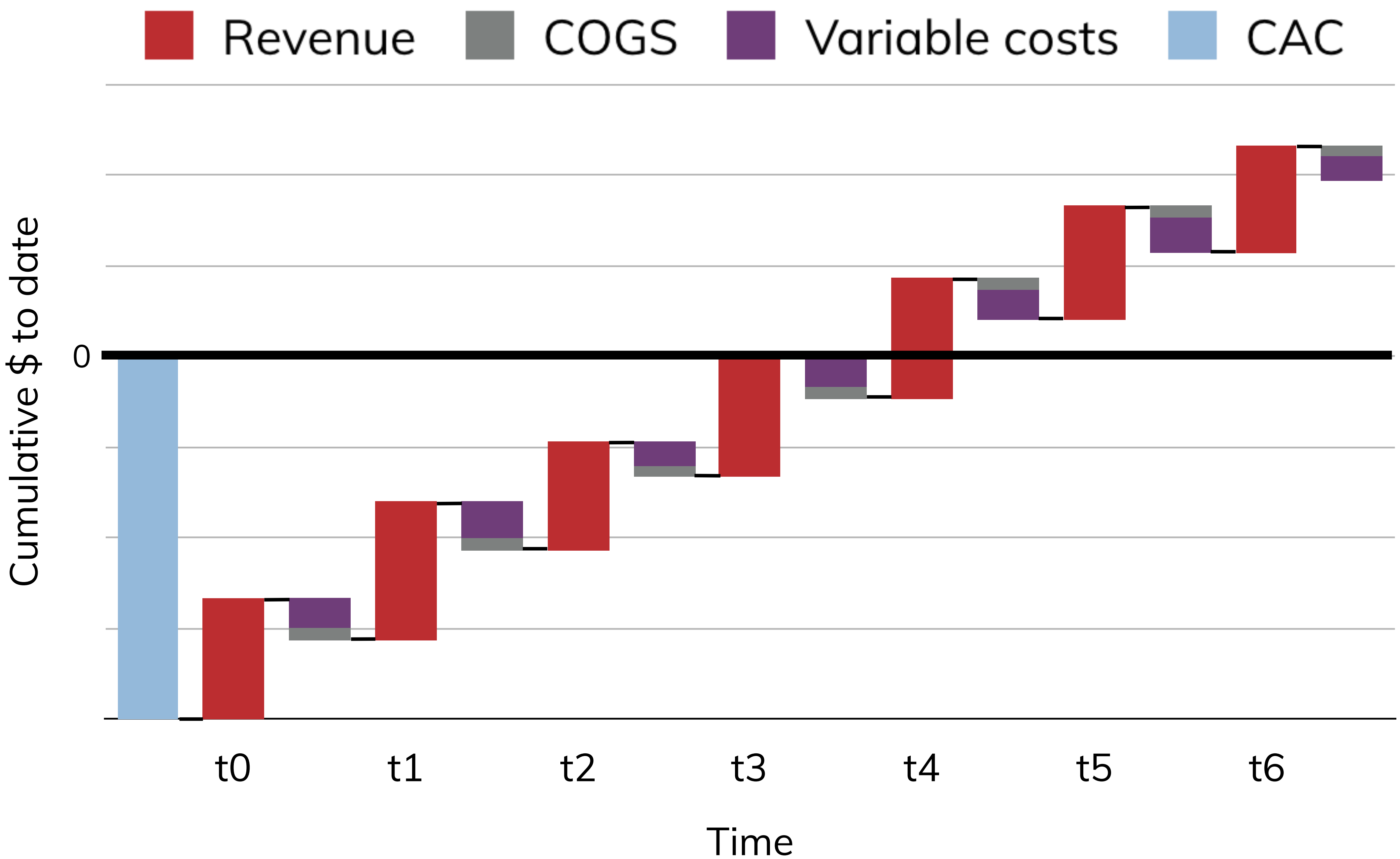
You start to think about all the accounting cash flows associated with that customer...

- Purchases (Revenue)
- Cost of goods sold (COGS)
- Variable costs
- Customer acquisition cost (CAC)

One way of looking at this...



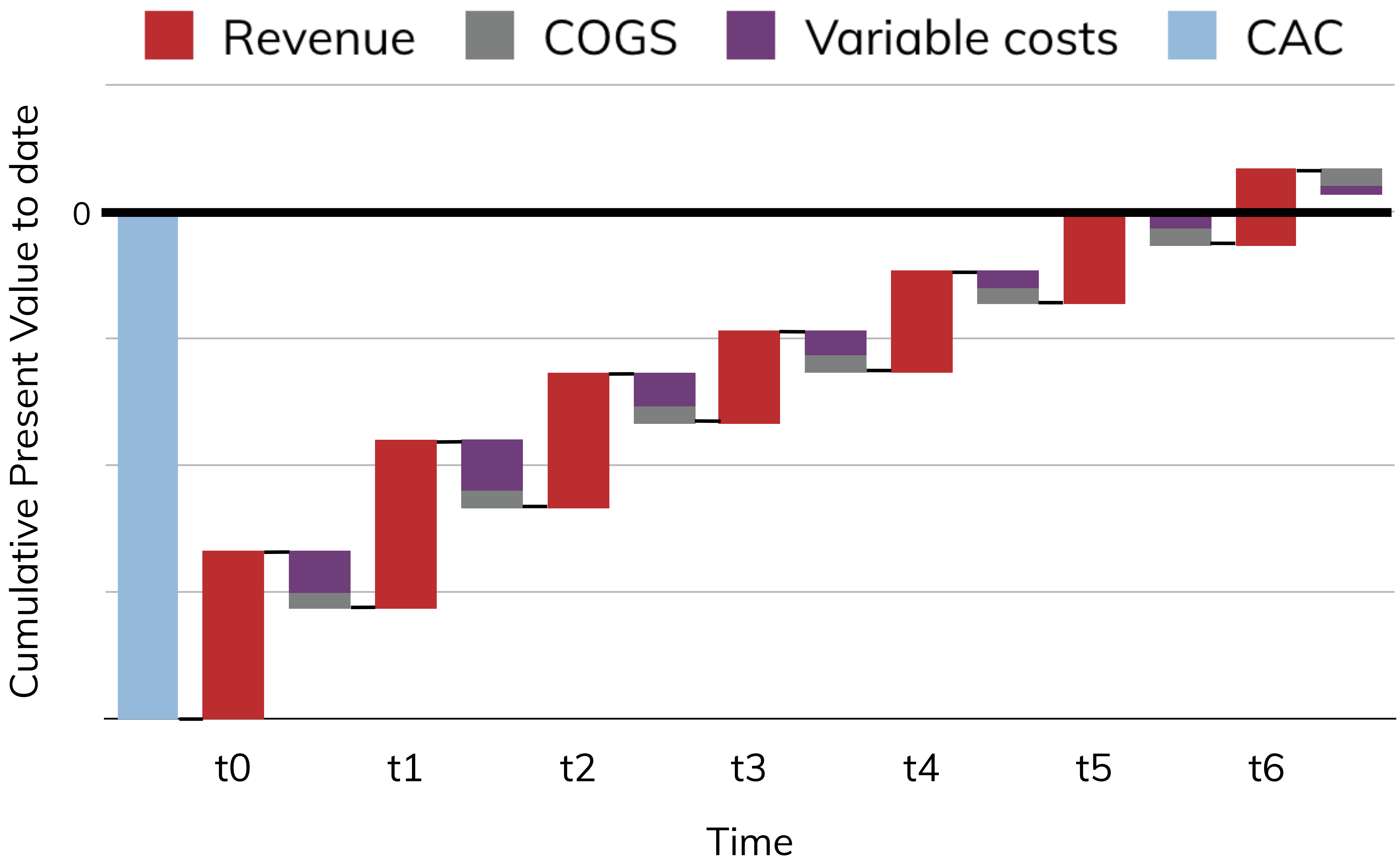
… replotted for shape …



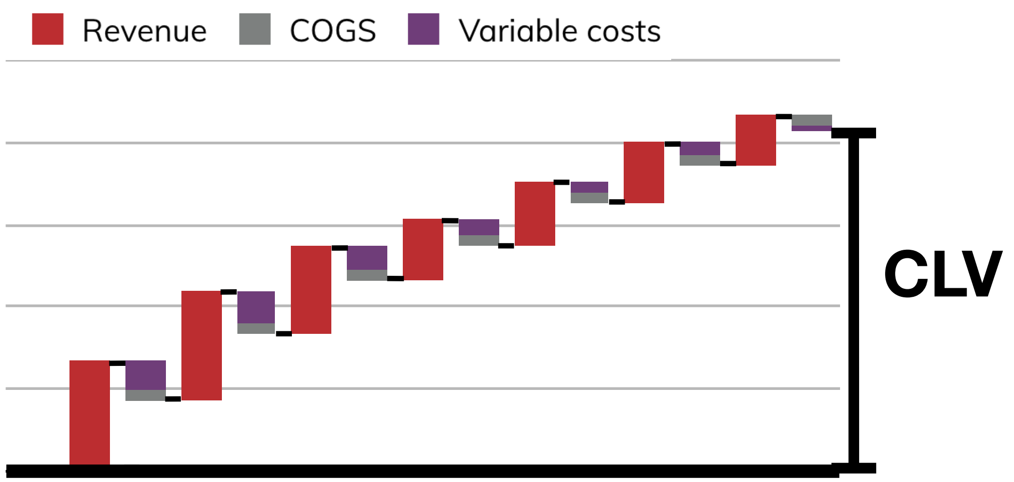
Let’s define “Customer Lifetime Value”

Net Present Value

$$PV = \frac{\$}{(1 + \text{discount\_rate})^{n_{periods}}}$$



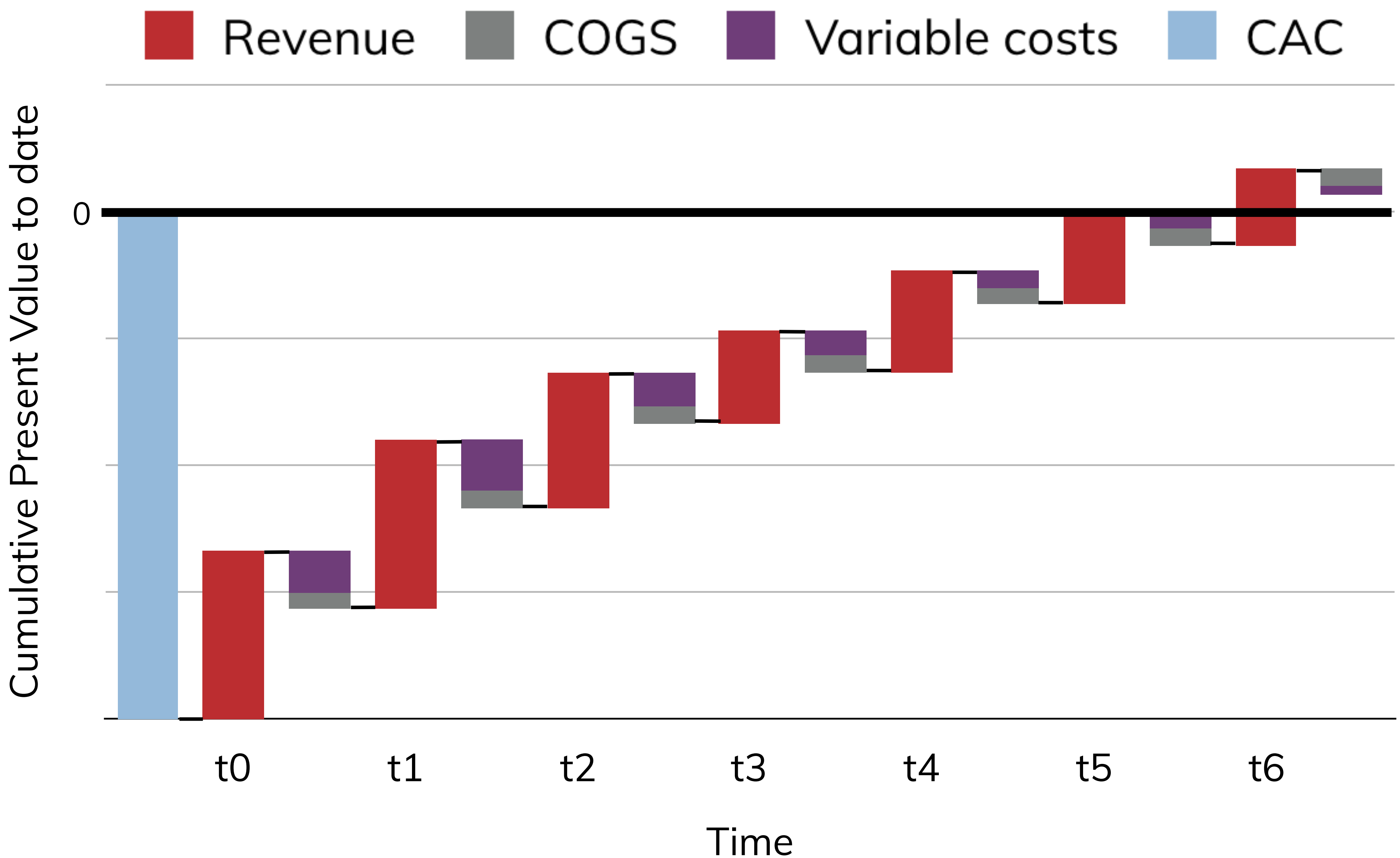
CLV typically defined as:



- **Revenue** over lifetime
- Minus **variable costs** (including costs of goods sold)
- **Discounted** at a company-specific discount rate
- [Does not include customer acquisition costs]

Whatever you choose, please define your metrics.

90% of business analysis is reasoning about this



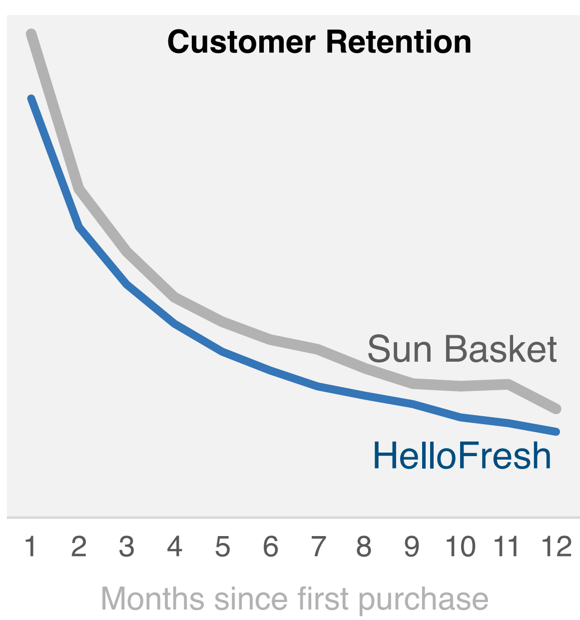
II. Exploration

But first, about us:

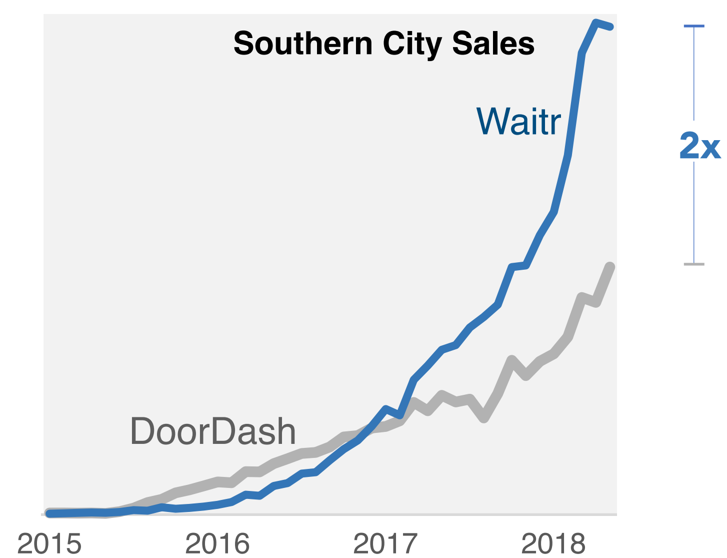
Second Measure analyzes billions of credit card transactions to answer real-time questions about consumer behavior

We answer questions like...

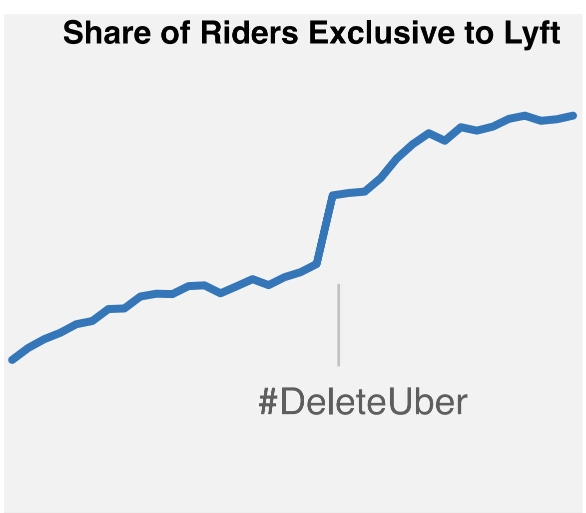
How well is Hello Fresh retaining its customers?



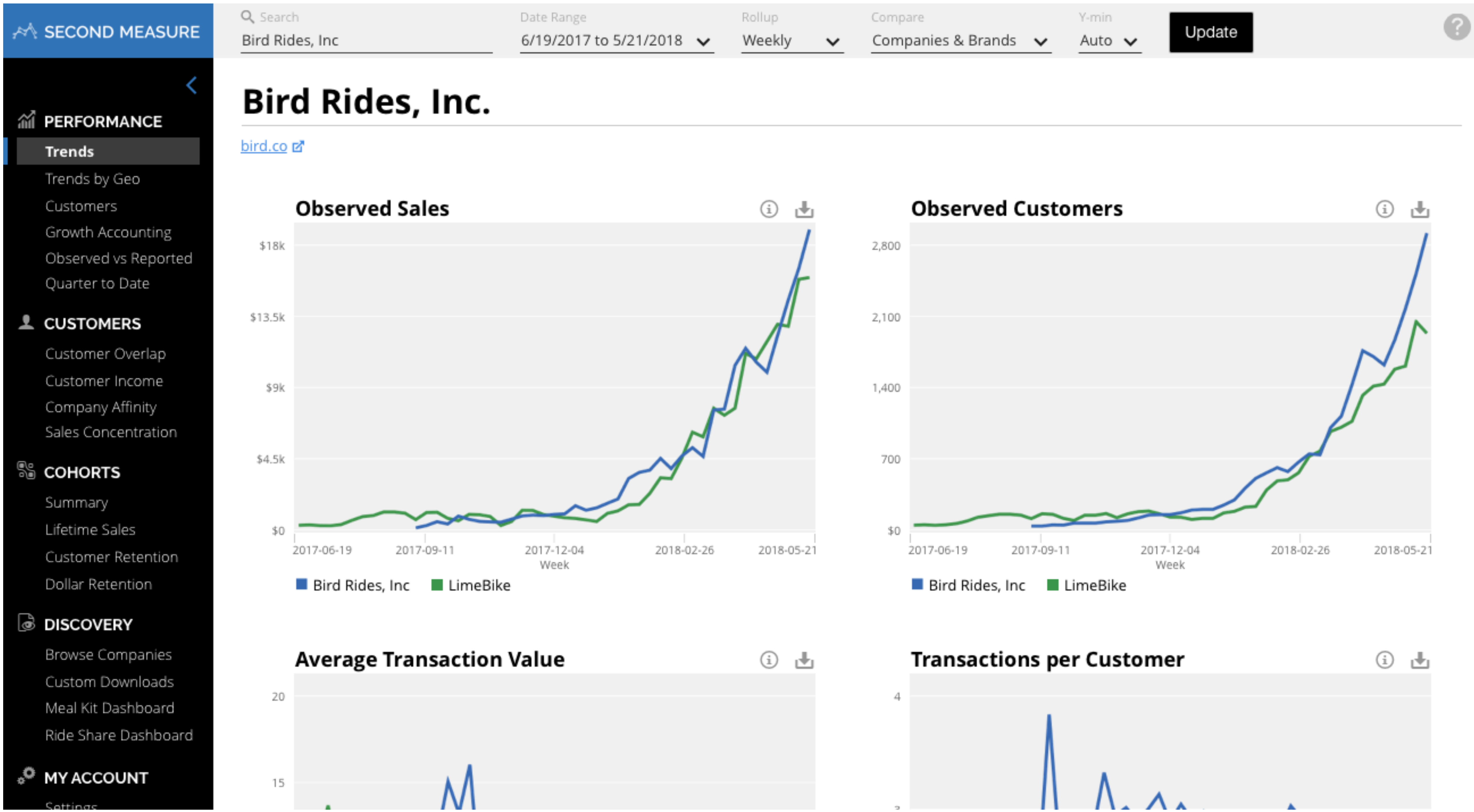
Is Waitr overtaking regional competitors?



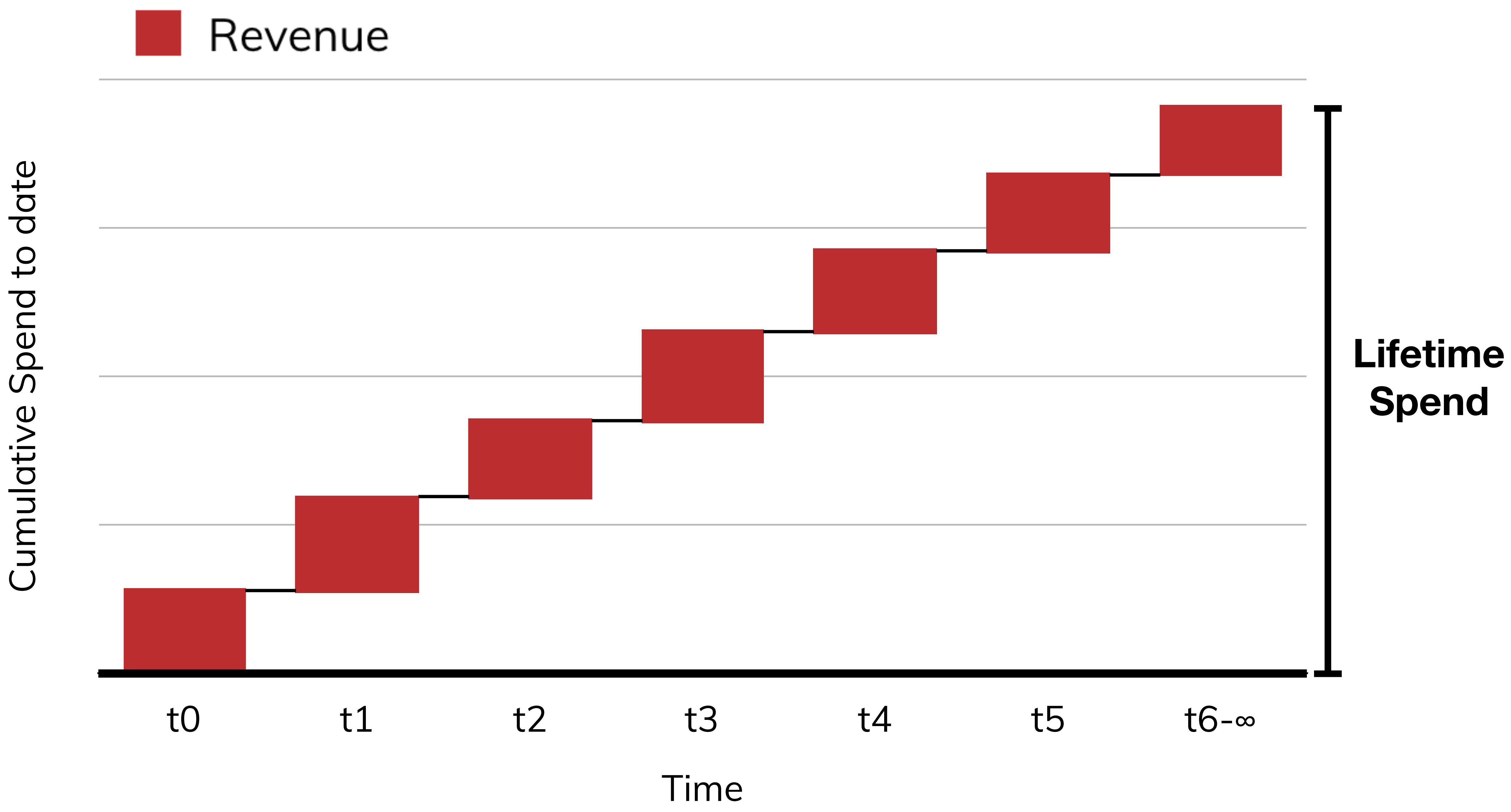
Did Lyft benefit from #DeleteUber?



With a self-service analytics platform



Confine ourselves to: lifetime spend



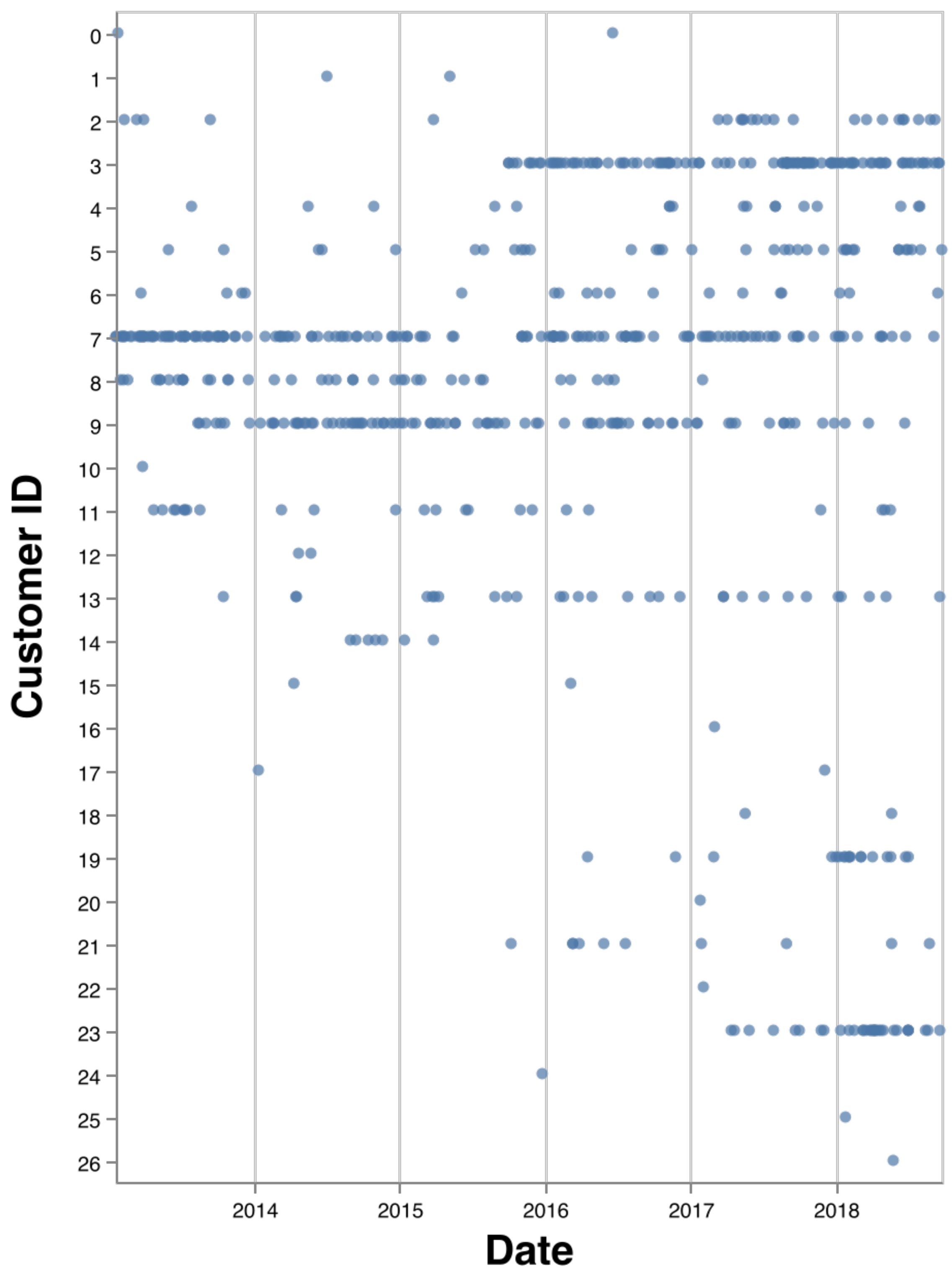
Confine ourselves to: lifetime spend

- Gross cumulative **sales** (including taxes)
- **Undiscounted**
- Nothing netted out

Case study to begin with:
Dollar General

- “an American chain of variety stores headquartered in Goodlettsville, Tennessee”
- The company that “went where they ain’t”
- Won a bidding war against Dolce and Gabbana for http://dg.com/ 💰

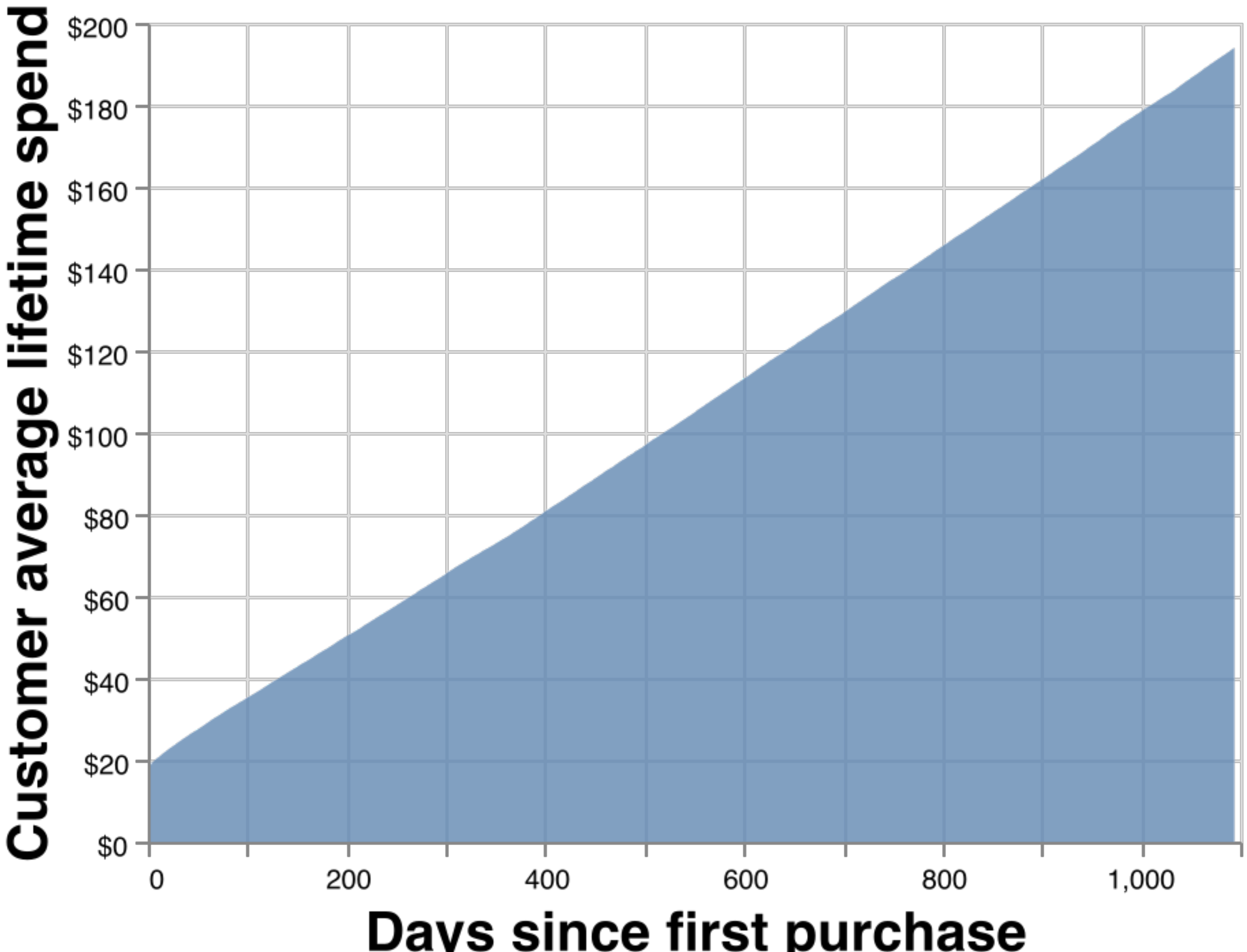
Raw purchases for Dollar General

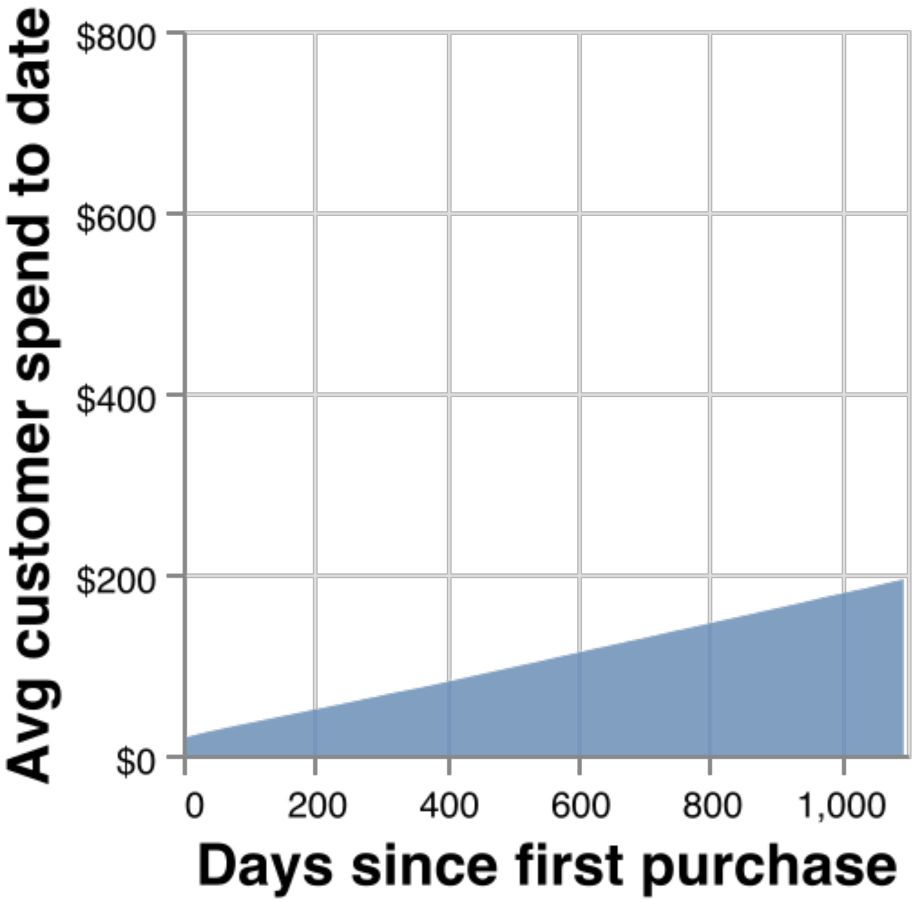


Our first calculation ("lifetime spend"), defined:

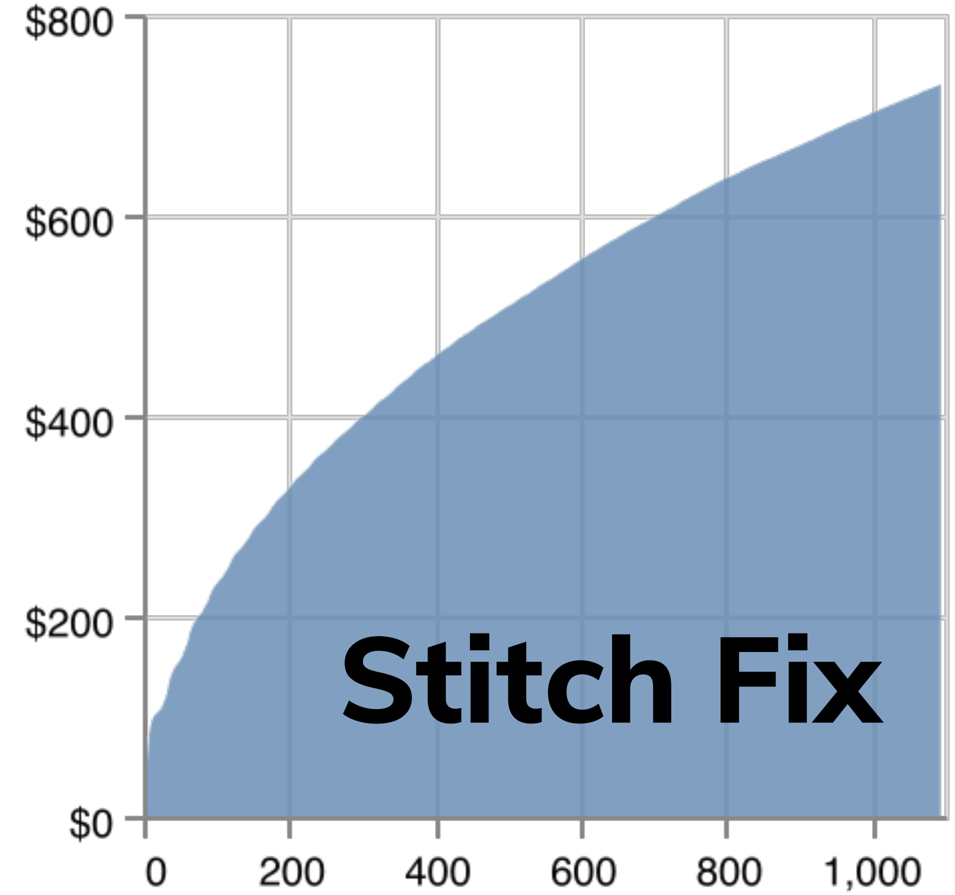
- Align all customers by the date of first purchase
- Calculate cumulative spending to date over time
- Average across all customers

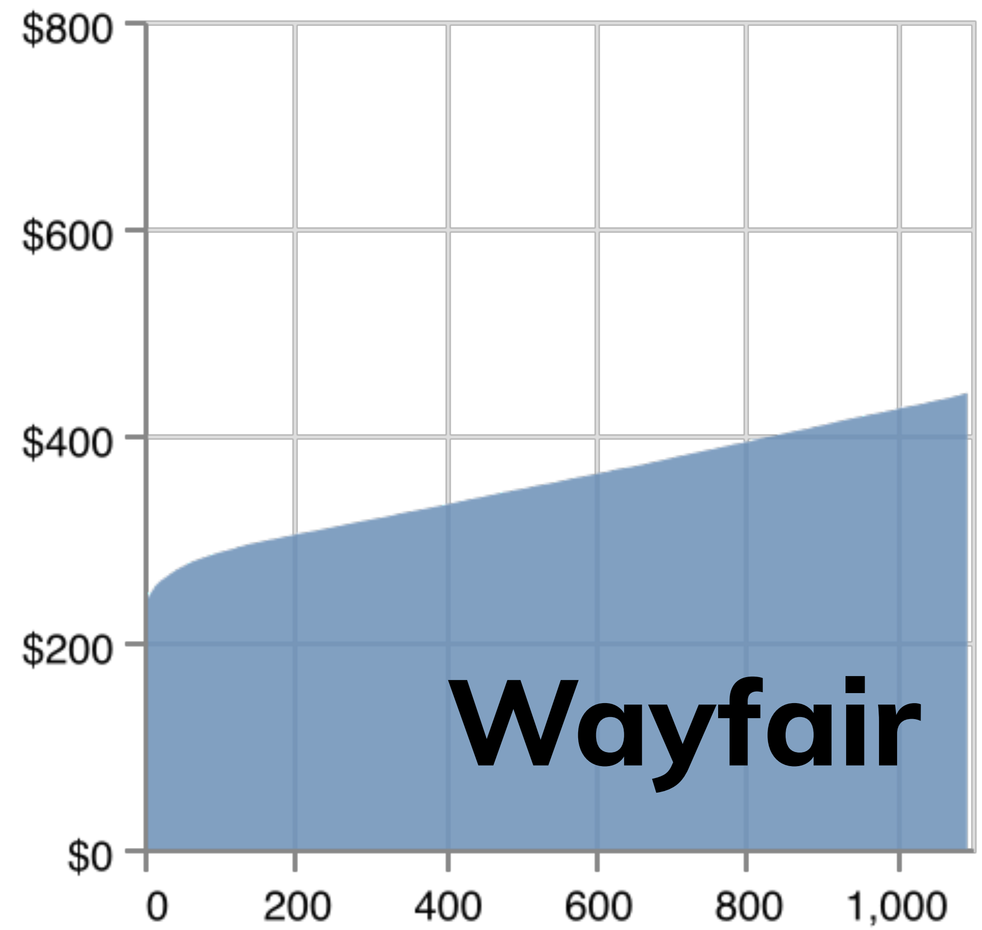
Dollar General Lifetime Spend



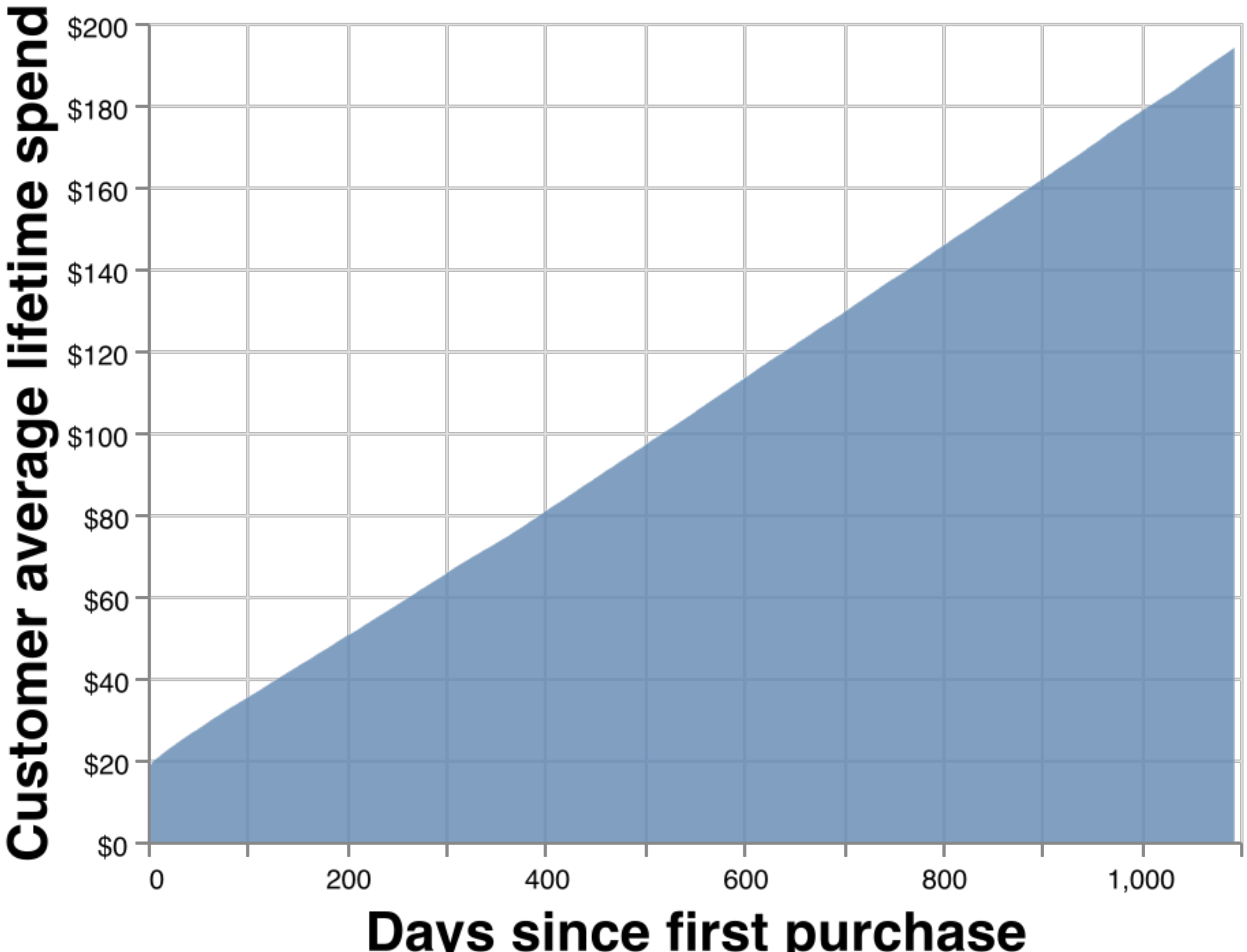


Dollar General



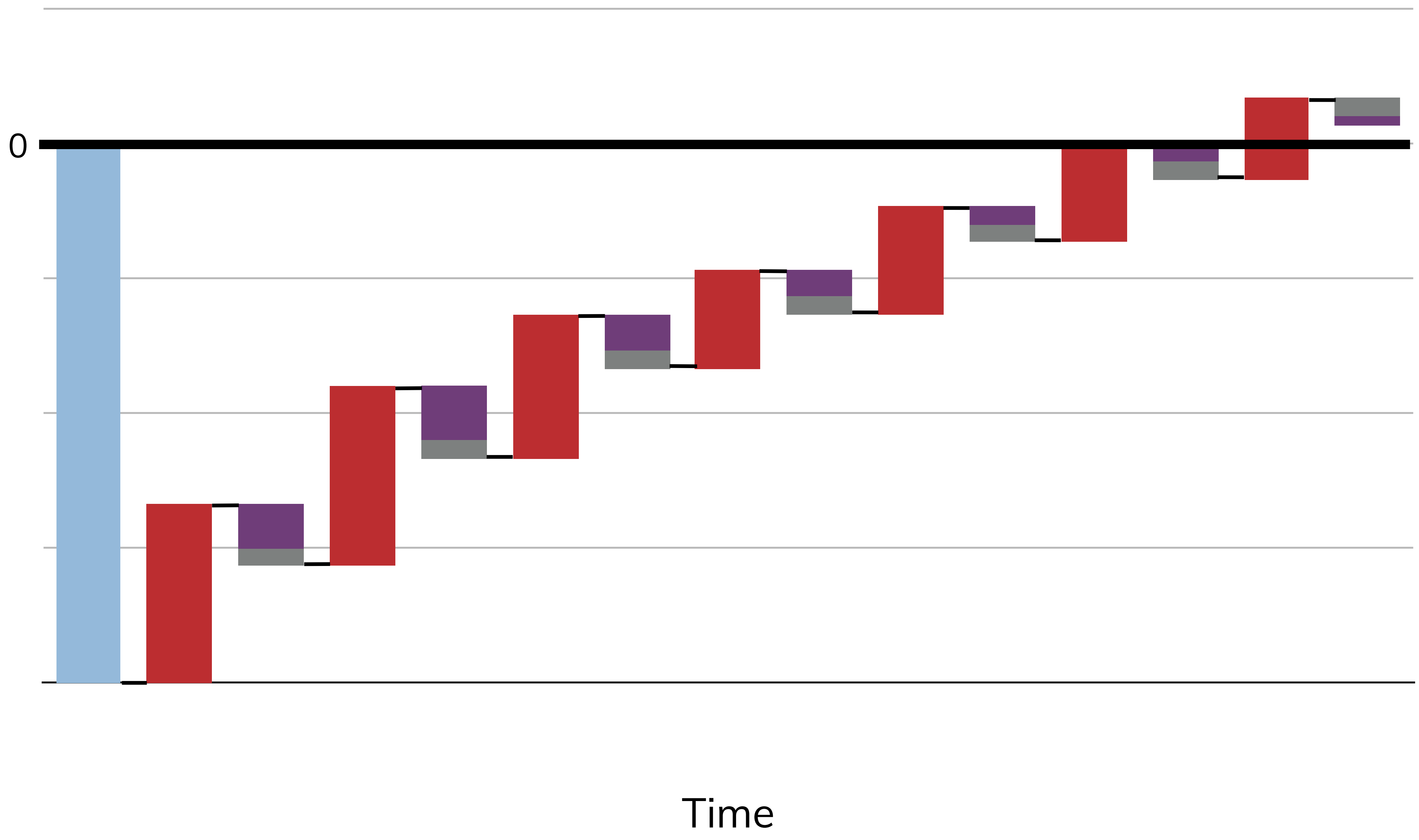


What's unsatisfying about this picture?

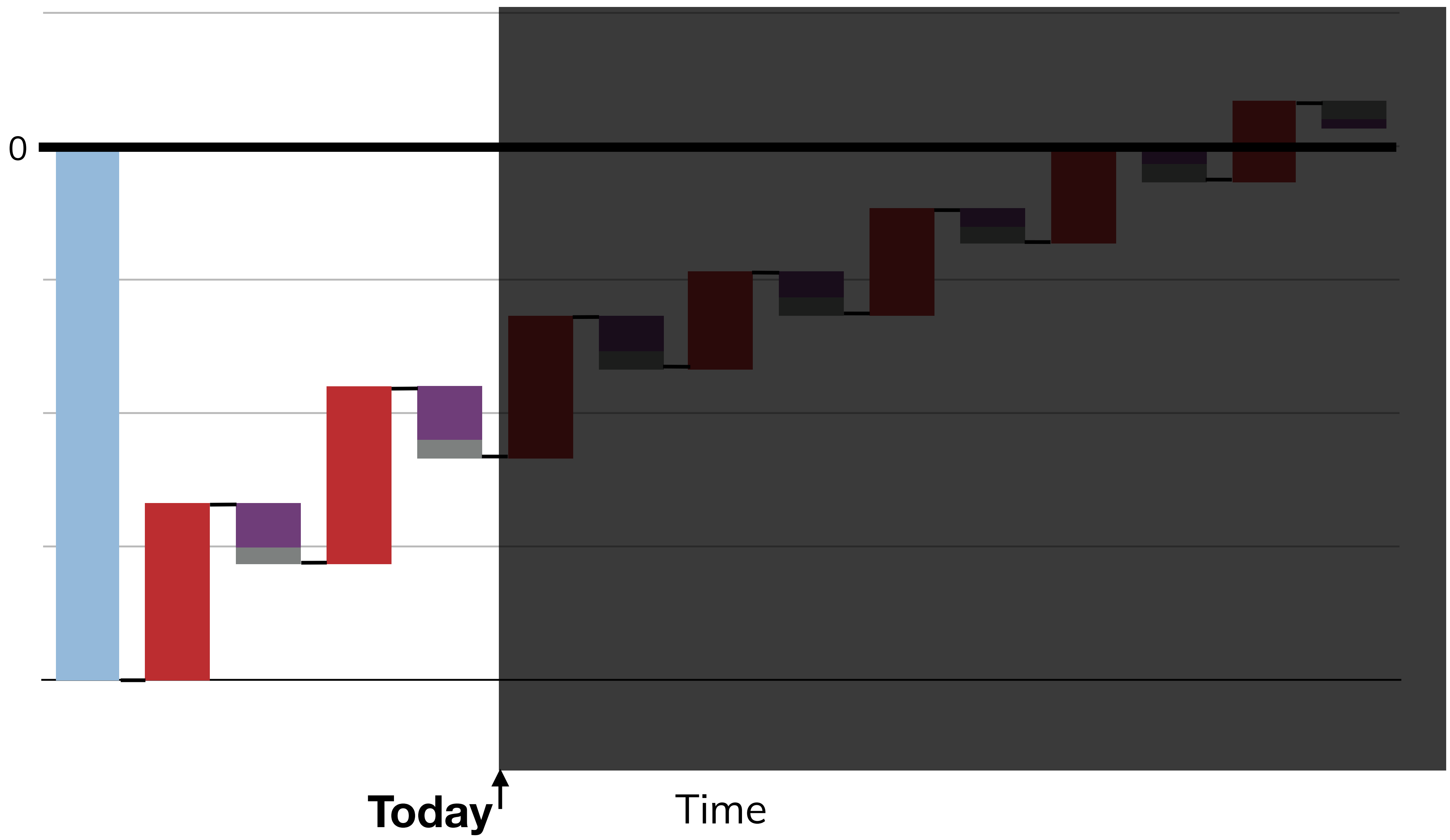


“Lifetime spend”, a hidden caveat

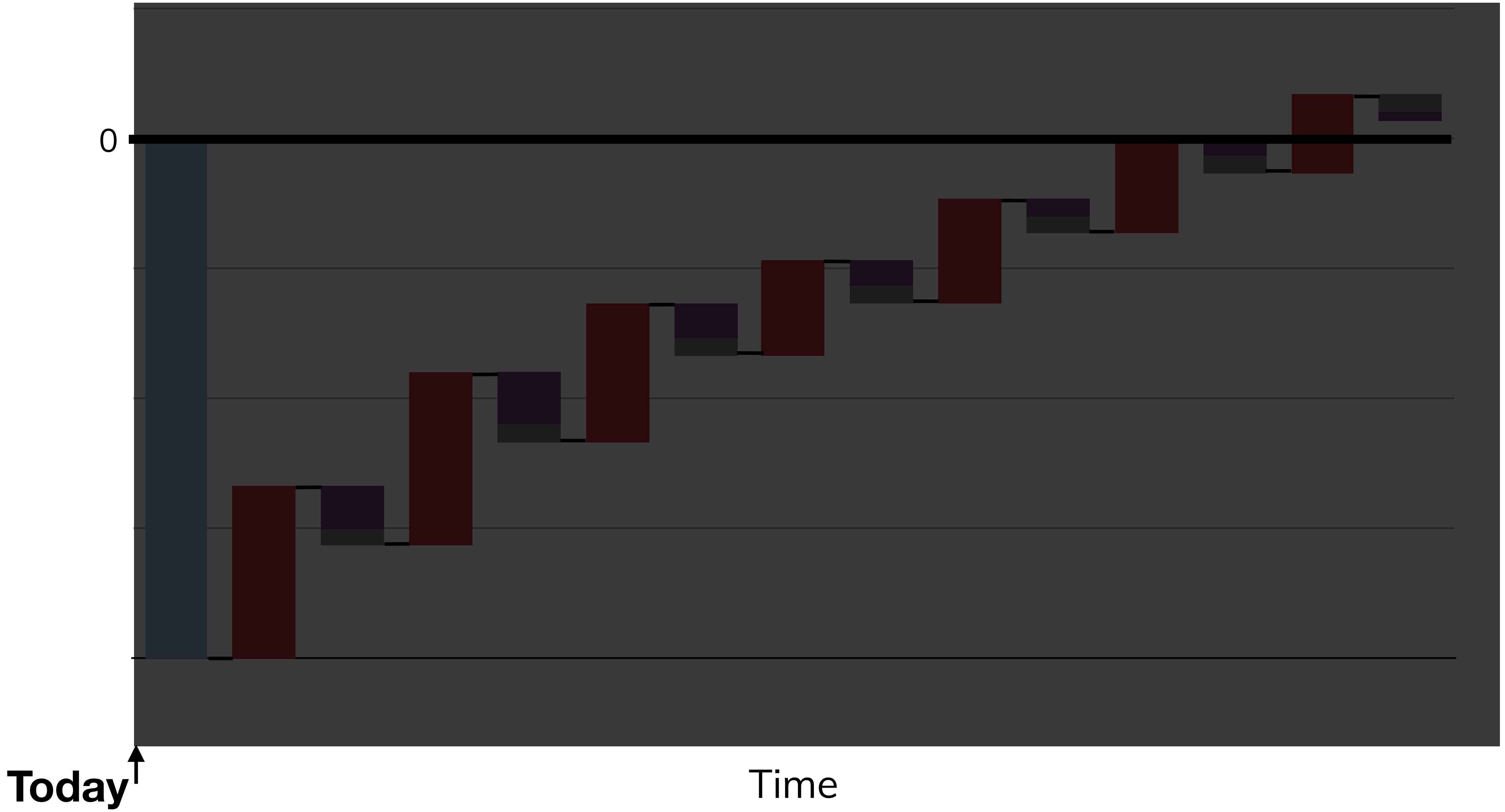
- Align all customers by the date of first purchase
- *Exclude customers who don’t have enough history (3 years)*
- Calculate cumulative spending to date over time
- Average across all customers


0
Time

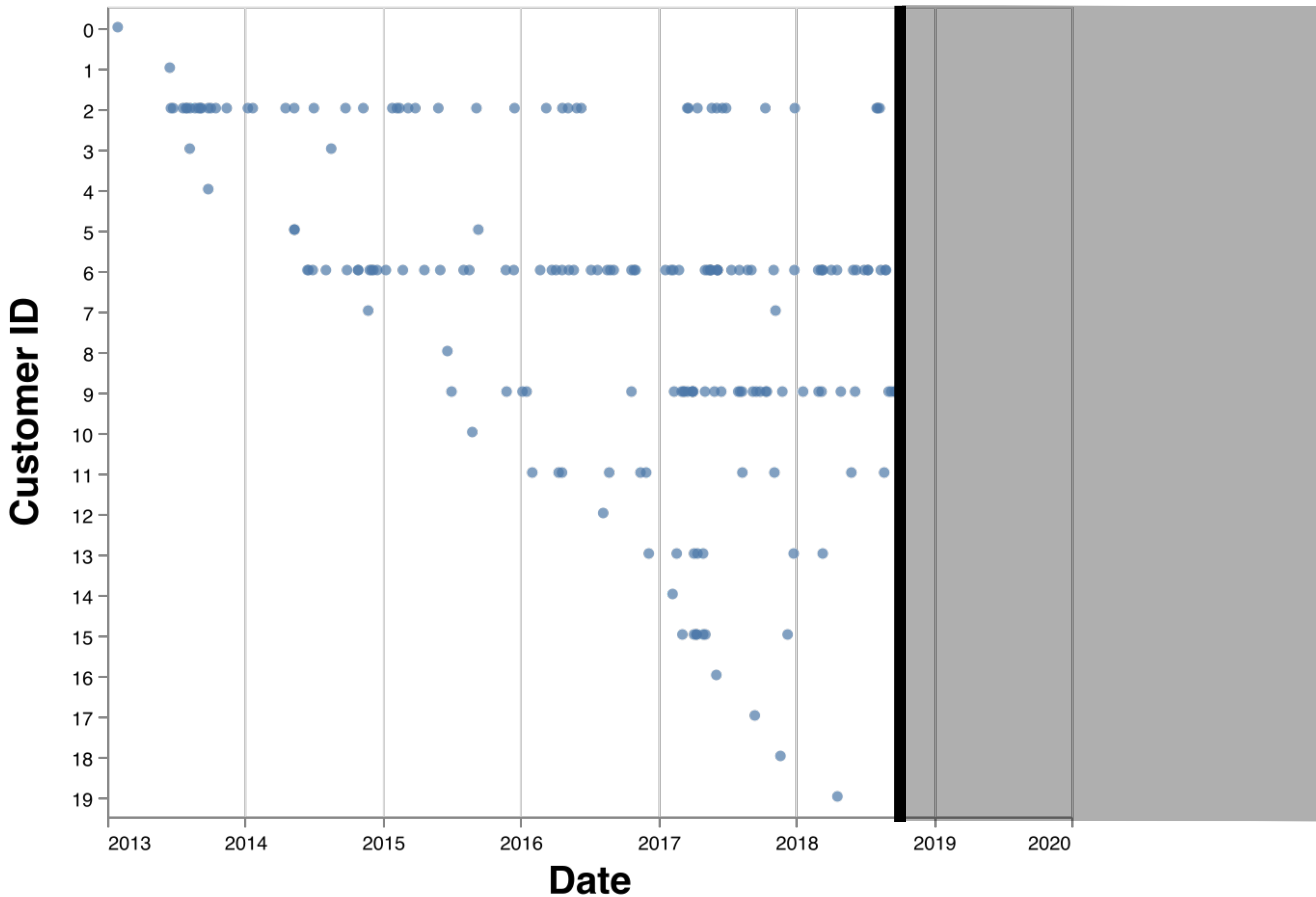
Most analysis:



Marketing analysis:



III. Estimation



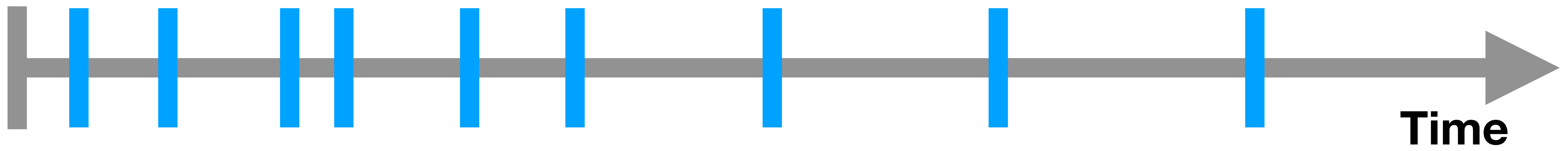
Customer ID
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
2013
2014
2015
2016
2017
2018
2019
2020
Date

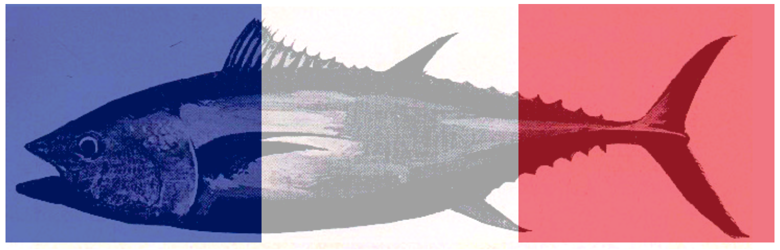
Factoring the problem

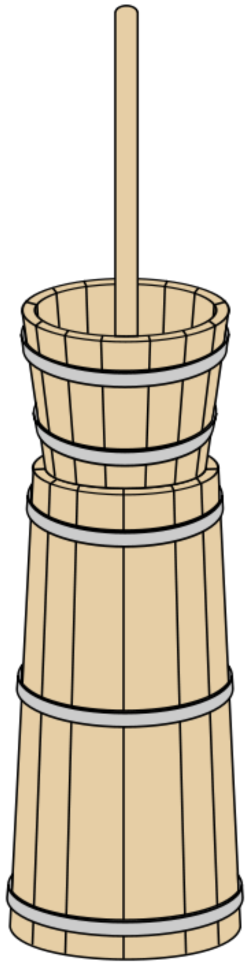
We’ll factor the estimation of customer purchase data into several steps. Let’s start by focusing on modeling transaction counts.

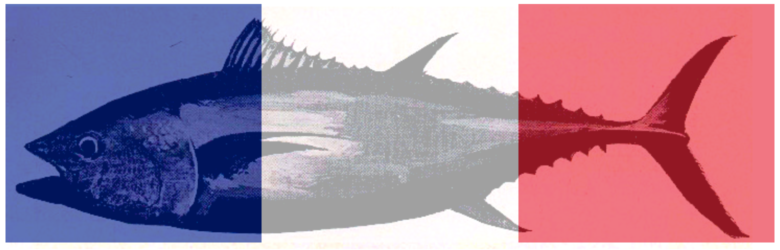
What’s the simplest thing that could work?

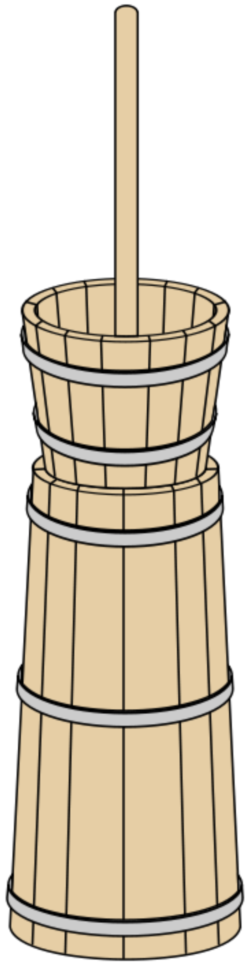
Answer: A Poisson process







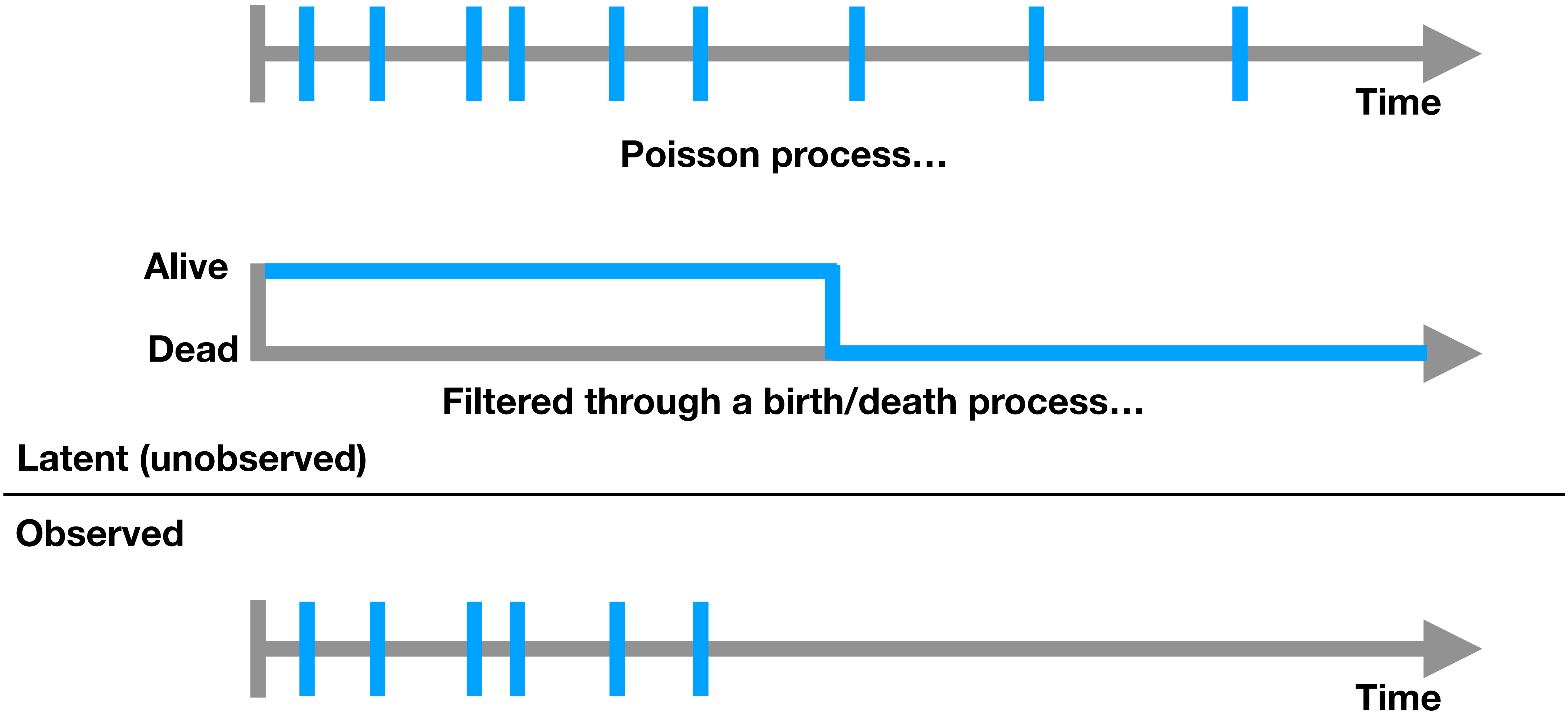




Poisson **Don’t** **Churn**

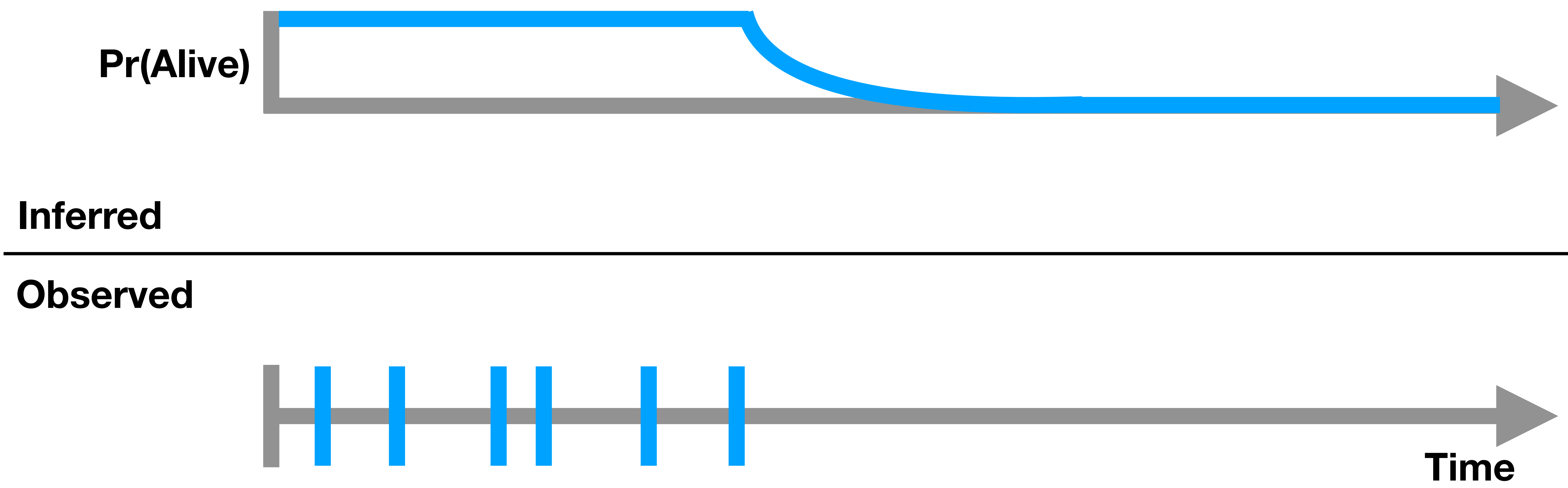
Ok then, what’s the next simplest thing that could work?

Our assumed data generating process:



Infer parameters given likelihood:

1. λ (i.e. rate parameter of Poisson distribution)
2. Hazard of churn per unit time



Factoring the problem

In this framework, we factor customer estimation into:

- Transaction count
- Churn date

Modeled jointly. Then, separately:

- Transaction amounts
- Costs
- ...

Are we sure this is simplest thing that might work?

Customer ID	Period	Transaction count
1	2018-01-01	1
1	2018-01-03	3
1	2018-01-06	2

Customer ID	Period	Transaction count
1	2018-01-01	1
1	2018-01-03	3
1	2018-01-06	2

$\bar{\lambda}$ = Mean(txn count) = 2

Customer ID	Period	Transaction count
1	2018-01-01	1
1	2018-01-02	0
1	2018-01-03	3
1	2018-01-04	0
1	2018-01-05	0
1	2018-01-06	2

Customer ID	Period	Transaction count
1	2018-01-01	1
1	2018-01-02	0
1	2018-01-03	3
1	2018-01-04	0
1	2018-01-05	0
1	2018-01-06	2

λ = Mean(txn count) = 1

Customer ID	Period	Transaction count
1	2018-01-01	1
1	2018-01-02	0
1	2018-01-03	3
1	2018-01-04	0
1	2018-01-05	0
1	2018-01-06	2
1	2018-01-07	0
1	2018-01-08	0

Customer ID	Period	Transaction count
1	2018-01-01	1
1	2018-01-02	0
1	2018-01-03	3
1	2018-01-04	0
1	2018-01-05	0
1	2018-01-06	2
1	2018-01-07	0
1	2018-01-08	0

How many rows of “negative space” should we account for after the last purchase?

Simulation results: bias by rate

True rate (events per unit time)	Percentage bias
0.1	158%
0.2	63%
0.5	19%
1	6.4%
2	1.6%
5	0.19%
10	0.01%

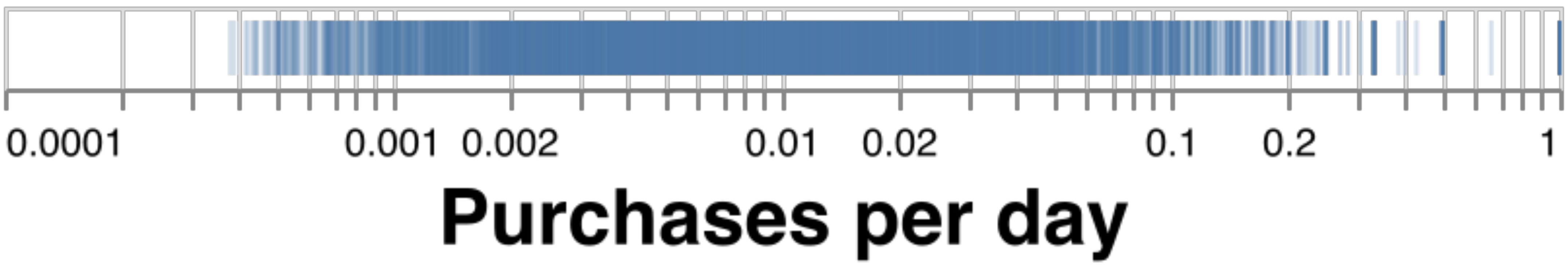
The first big idea: a hybrid probabilistic model

The second big idea: customer heterogeneity

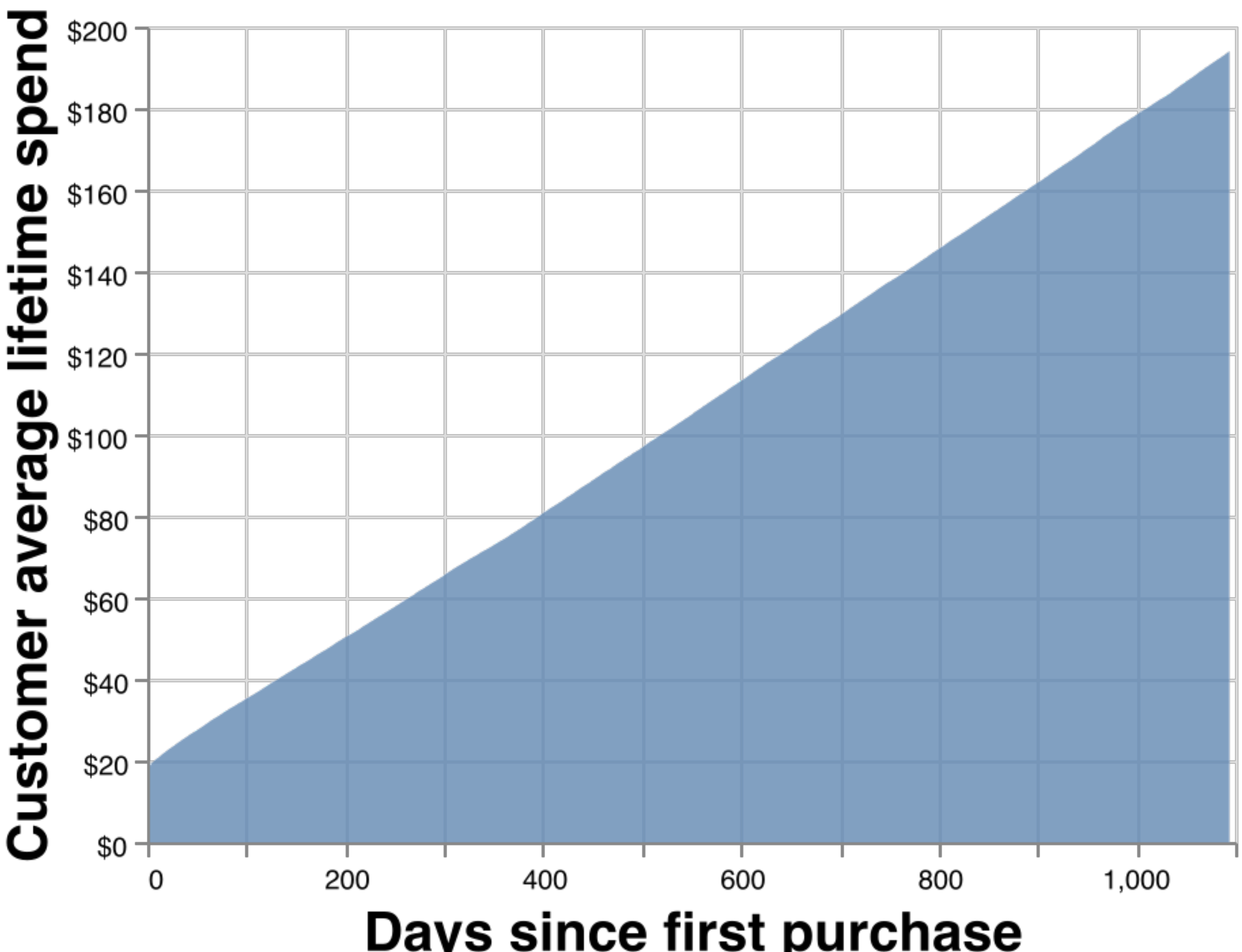
Purchases per day

- Observational unit: one customer
- Calculation: $\dfrac{\text{\# purchases}}{\max(\text{date}) - \min(\text{date})}$

Purchases per day (for Dollar General)

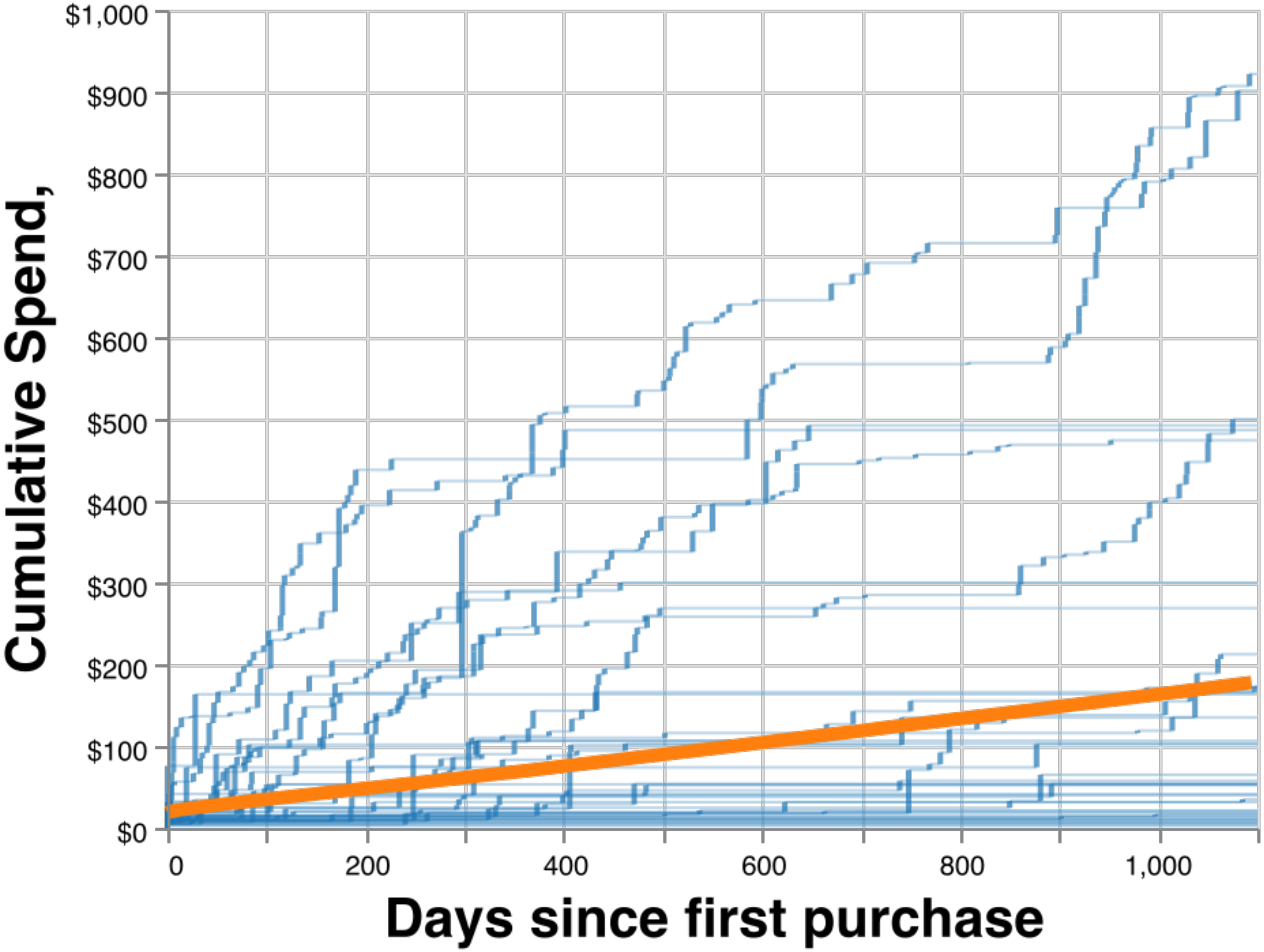


- Observational unit: one customer
- Calculation: $\dfrac{\#\ \text{purchases}}{\max(\text{date}) - \min(\text{date})}$

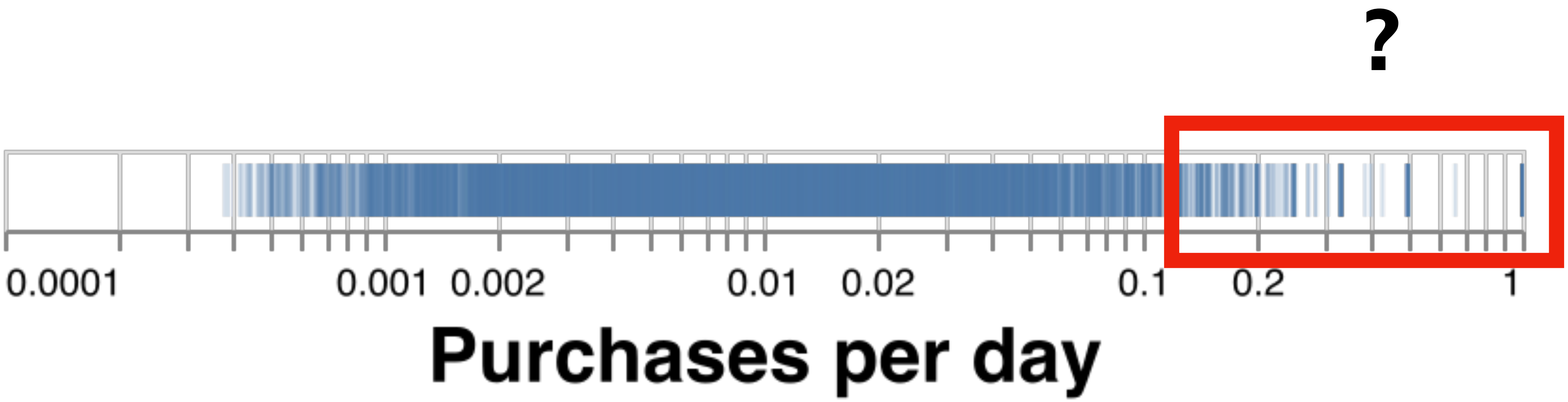


Customer average lifetime spend
$200
$180
$160
$140
$120
$100
$80
$60
$40
$20
$0
0
200
400
600
800
1,000
Days since first purchase

Not so orderly now, huh?



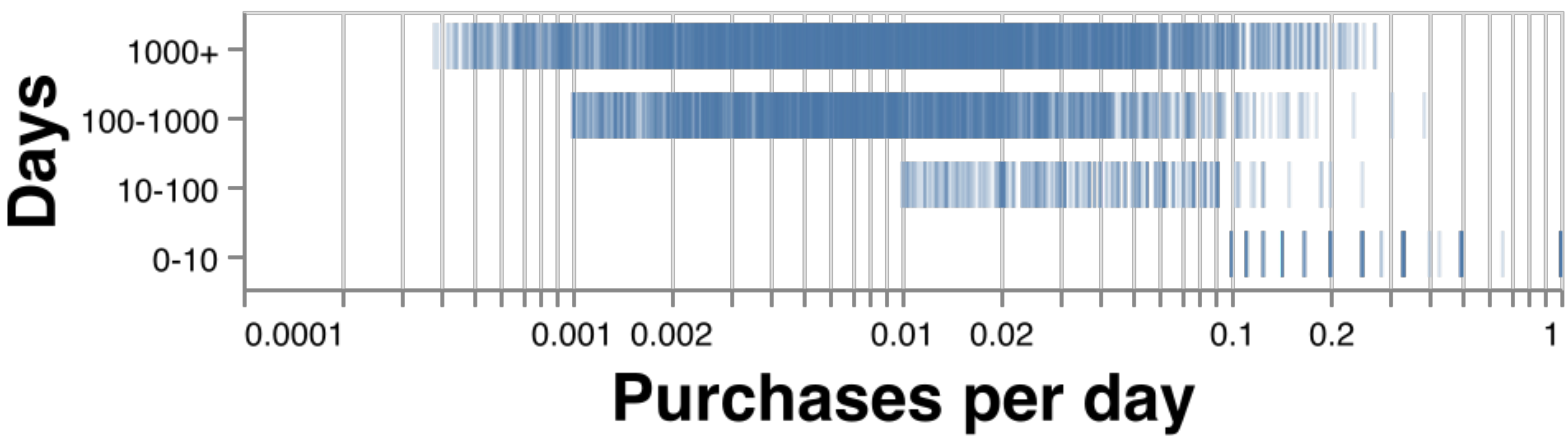
Blue = each customer, orange = average



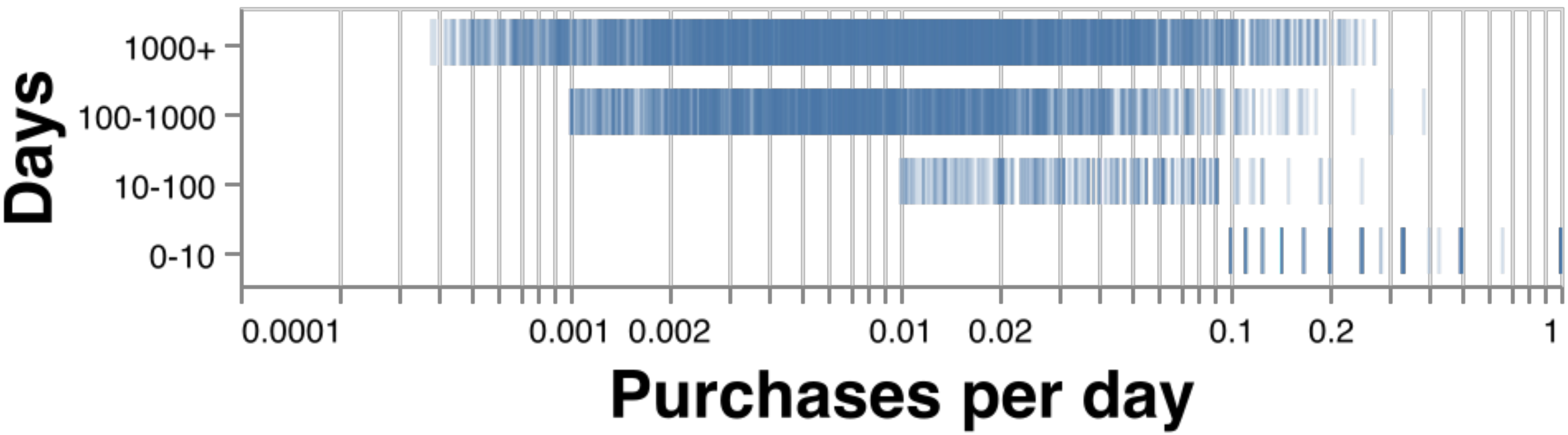
- Observational unit: one customer
- Calculation: $\dfrac{\#\ \text{purchases}}{\max(\text{date}) - \min(\text{date})}$

- Calculation: $\dfrac{\#\ \text{purchases}}{\max(\text{date}) - \min(\text{date})}$

Split by the denominator:



A solution: multilevel models



Jointly Estimate customer rate and a distribution over all those rates; distribution is fit from data and acts as a prior for rates with small N.

Estimate Your Lifetimes

https://github.com/CamDavidsonPilon/lifetimes

LIFETIMES

Measuring users is hard. Lifetimes makes it easy.

pypi package 0.9.1.0 | docs passing | build failing | coverage 97%

Introduction

Lifetimes can be used to analyze your users based on a few assumption:

1. Users interact with you when they are "alive".
2. Users under study may "die" after some period of time.

“Counting your Customers” Models in Lifetimes

Model	Paper	Comments
Pareto / Negative Binomial Distribution	Schmittlein, Morrison & Columbo 1987	The originator
Beta Geometric / Negative Binomial Distribution	Fader, Hardie & Lee 2005	Tweaked churn process that allows for a more efficient implementation; has a limitation in Pr(Alive) estimand
Modified Beta Geometric / Negative Binomial Distribution	Batislam, Denizel & Filiztekin 2007	Fixes limitation in BG/NBD’s Pr(Alive)

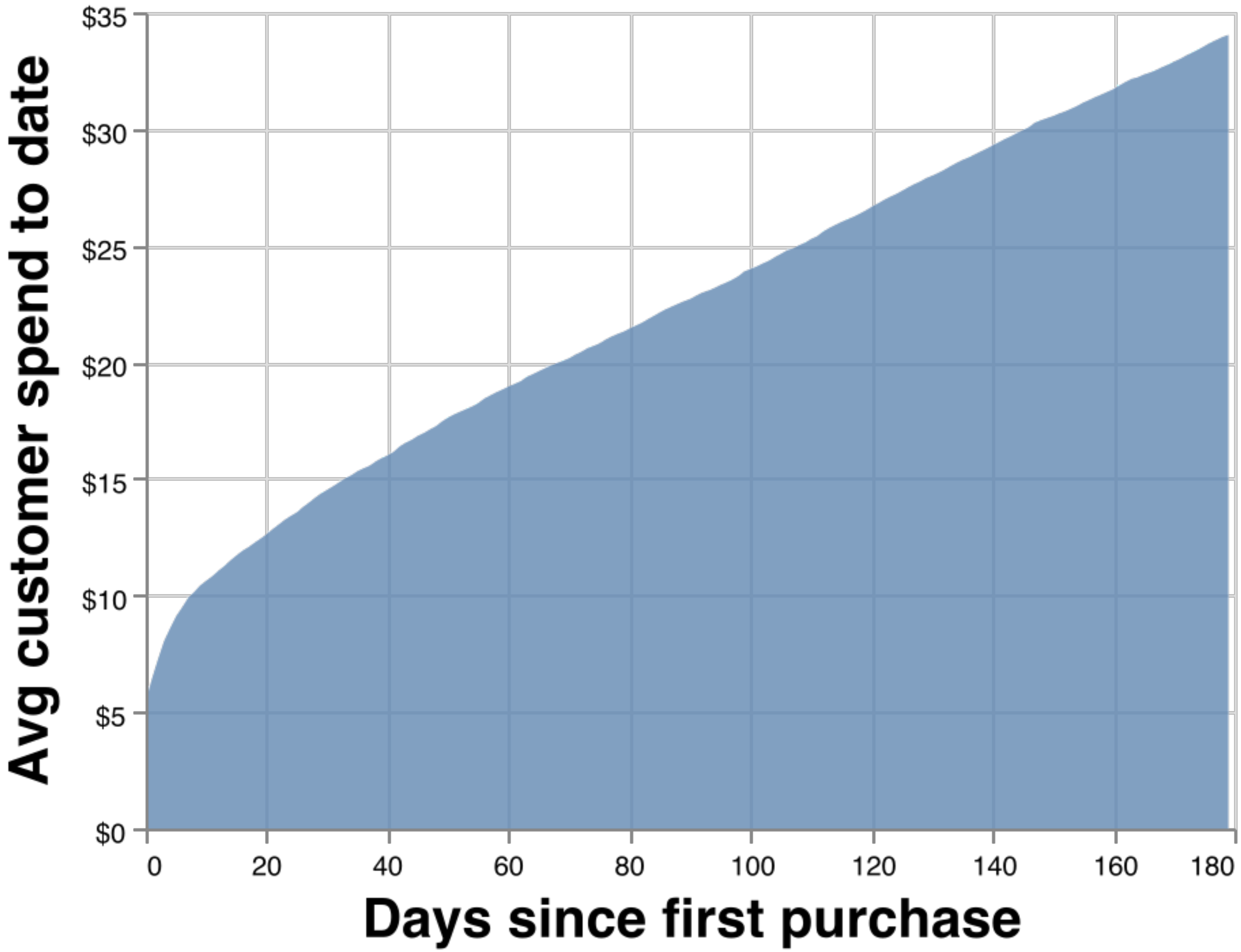
Recommendation: use this!

Model	Paper	Comments
Pareto / Negative Binomial Distribution	Schmittlein, Morrison & Columbo 1987	The originator
Beta Geometric / Negative Binomial Distribution	Fader, Hardie & Lee 2005	Tweaked churn process that allows for a more efficient implementation; has a limitation in Pr(Alive) estimand
Modified Beta Geometric / Negative Binomial Distribution	**Batislam, Denizel & Filiztekin 2007**	**Fixes limitation in BG/NBD's Pr(Alive)**

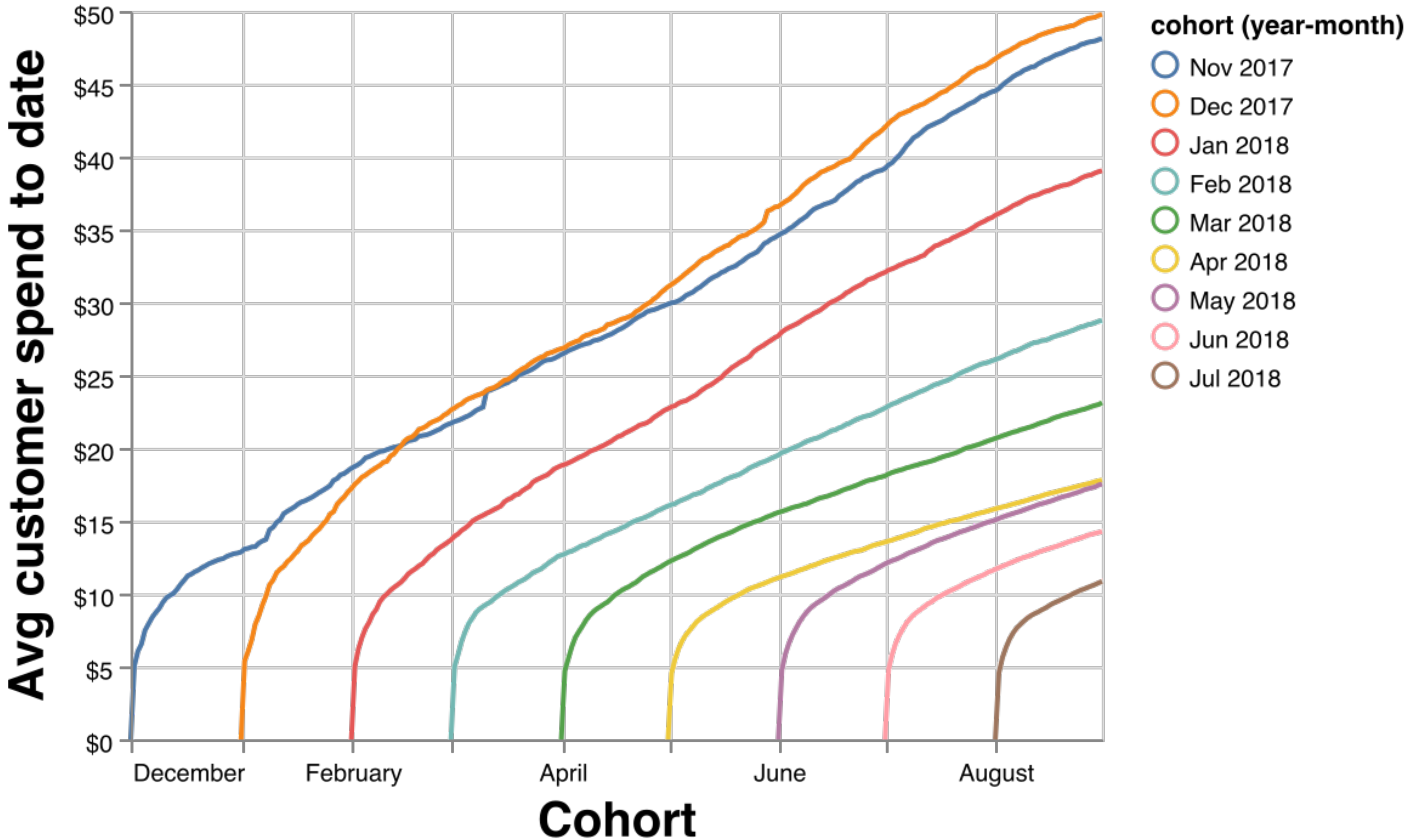
Next up: Bird Rides, Inc.

- Santa Monica-based scooter share startup and well known hockey-stick

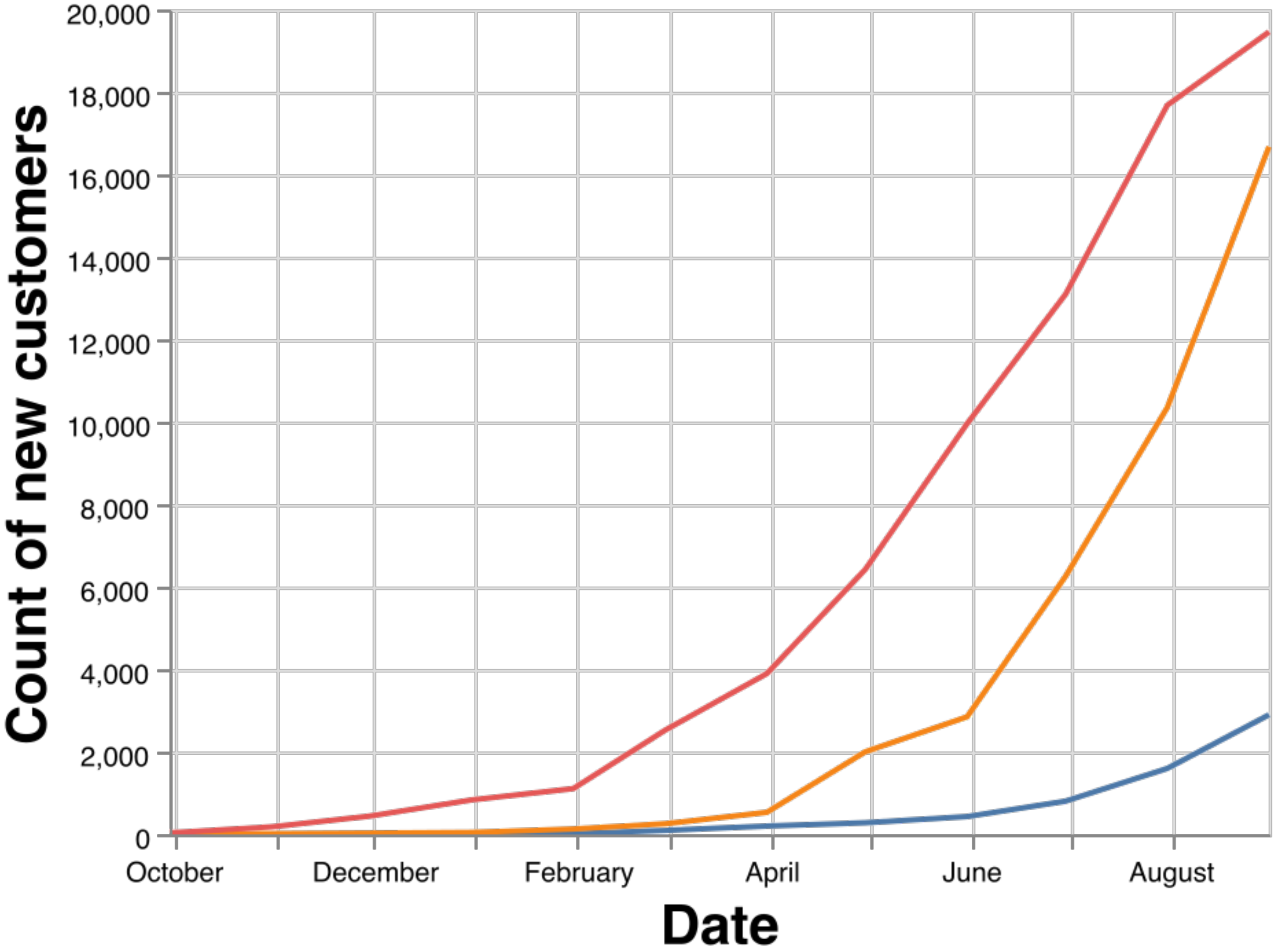
An unsatisfying lifetime spend chart

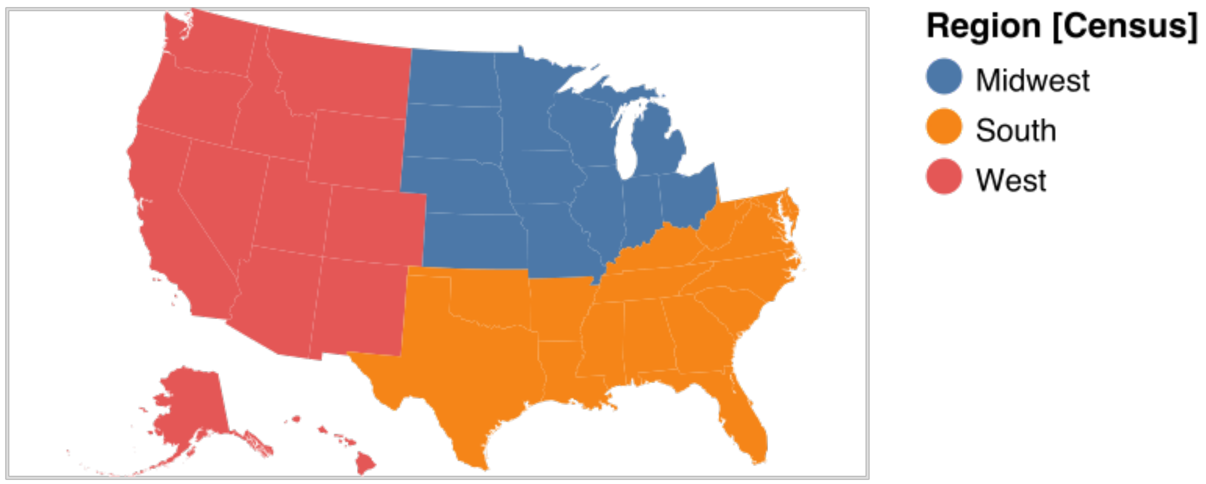


A band-aid: cohorted lifetime spend

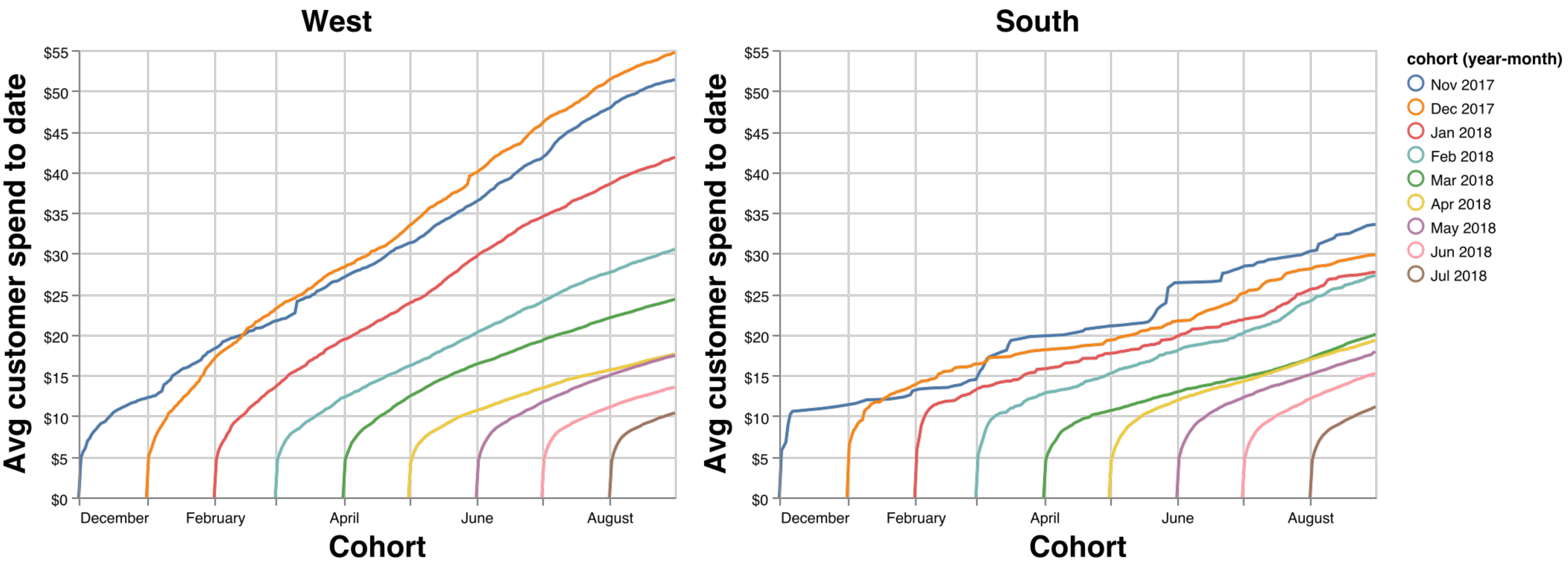


Bird: count of new customers by geo





Cohorted lifetime spend by geo



Bird: Model estimates

- 24 month expected total **sales** (including taxes)
- **Undiscounted**

Region	Estimated 24 month sales	% difference
West	$44	0%
South	$37	-16%
Midwest	$27	-37%

Wrap-up: the core ideas

- Workflow: build models from past data, project the near future. Analyze the 2 combined
- Probabilistic models can be composed of familiar building blocks to fit tailored situations
- Mind the noise!
- Customers vary widely
- Define your metrics, please!

Learn More

- Corp finance: Aswath Damodaran’s YouTube lectures
- Multilevel Models: Statistical Rethinking by McElraith
- Counting Your Customers: Lifetimes, Shopify blog
- Survival Analysis usecases: talk from Opendoor

Thank you!

- brian@secondmeasure.com
- Read http://blog.secondmeasure.com!
- We're hiring! In the bay area! Datasci, social scientists, data engineering & ETL, analysts

Questions?

- brian@secondmeasure.com
- Read http://blog.secondmeasure.com!
- We’re hiring! In the bay area! Datasci, social scientists, data engineering & ETL, analysts