Diving into the Data Science Job Market

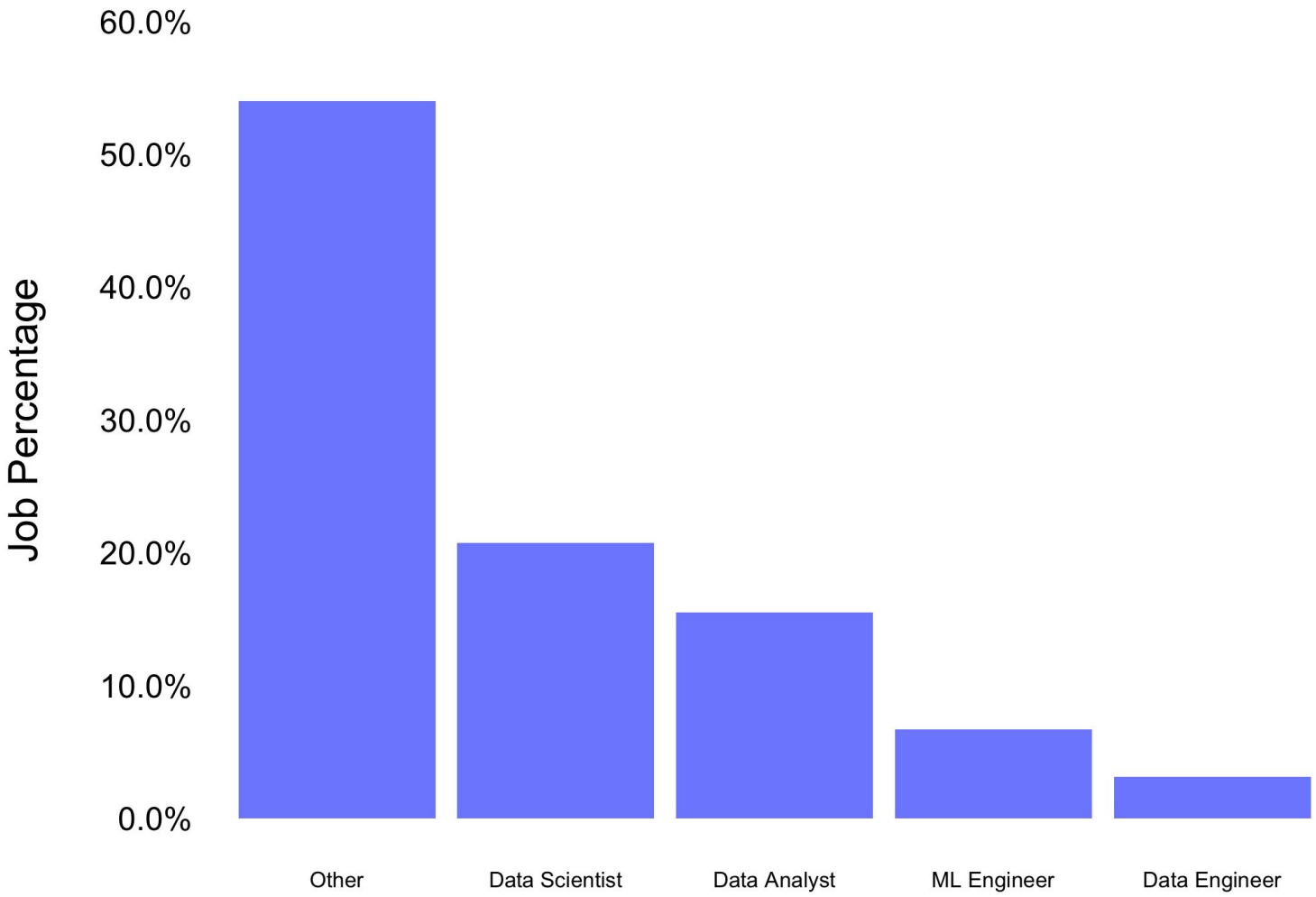
MSDS 593 Final Presentation

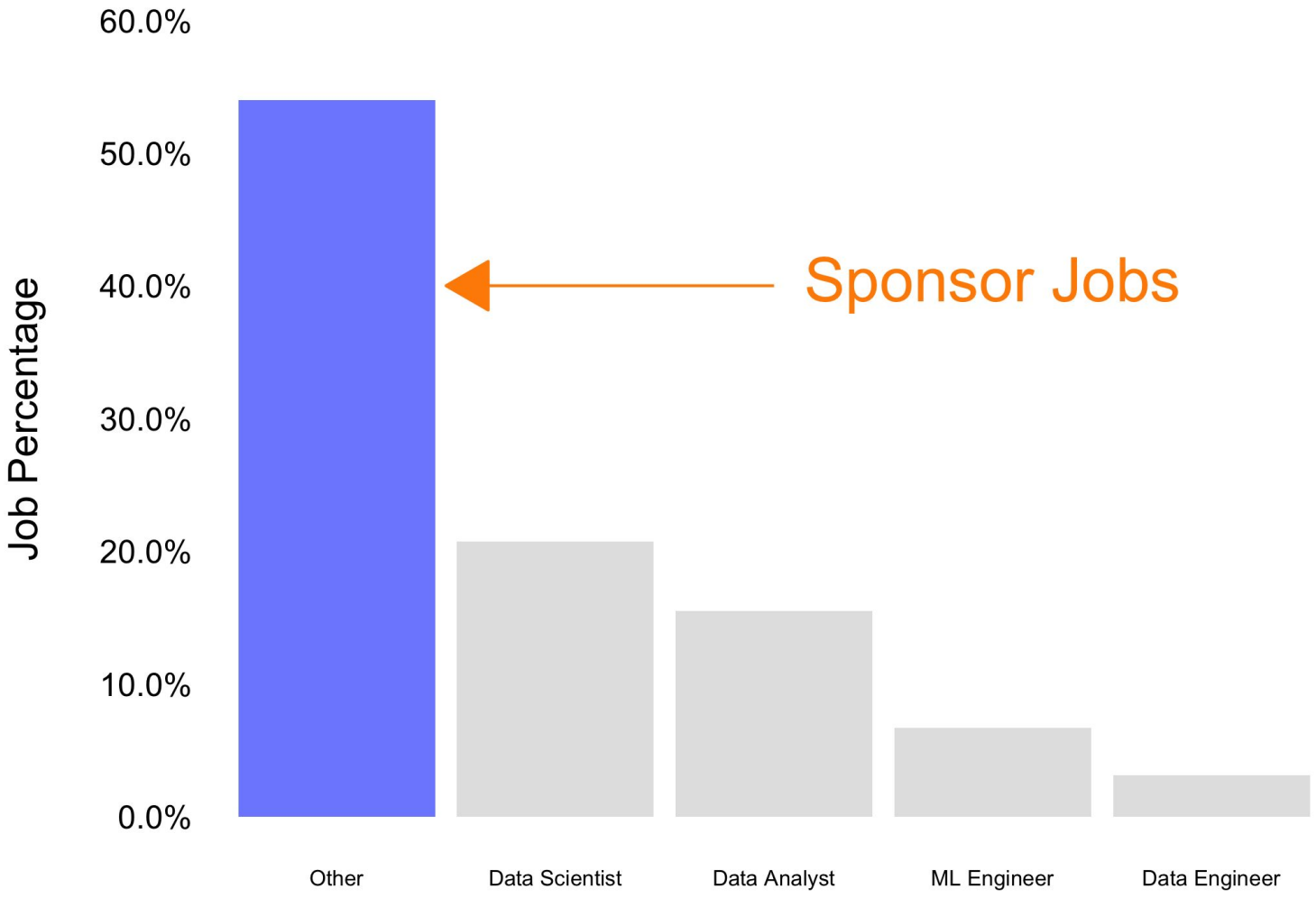
Shrikar Thodla
Dillon Quan

Datasets

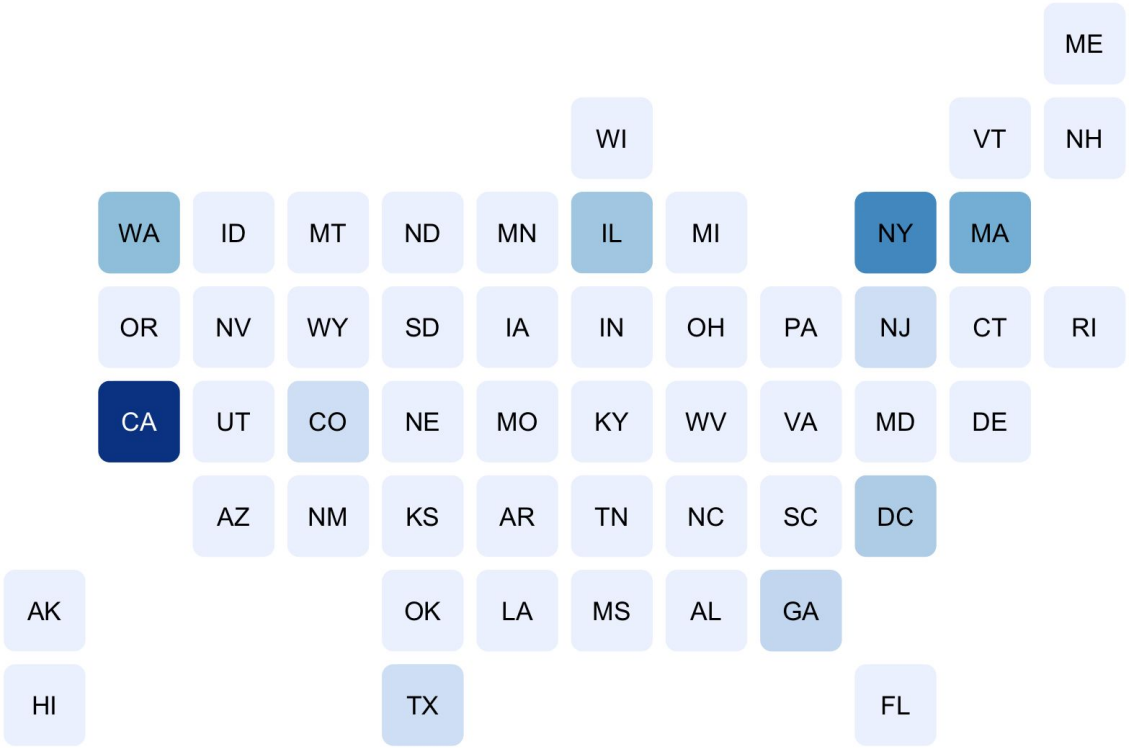
- Kaggle:
  - Data Science job posts on Indeed (2018) ~ 7000 samples
    - Features: Location, Job Title, Description
  - Most in demand skills for Data Science

- Web - Scraped:
  - Data Science job posts on Indeed (2019) ~ 24000 samples


Job Percentage
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Other
Data Scientist
Data Analyst
ML Engineer
Data Engineer

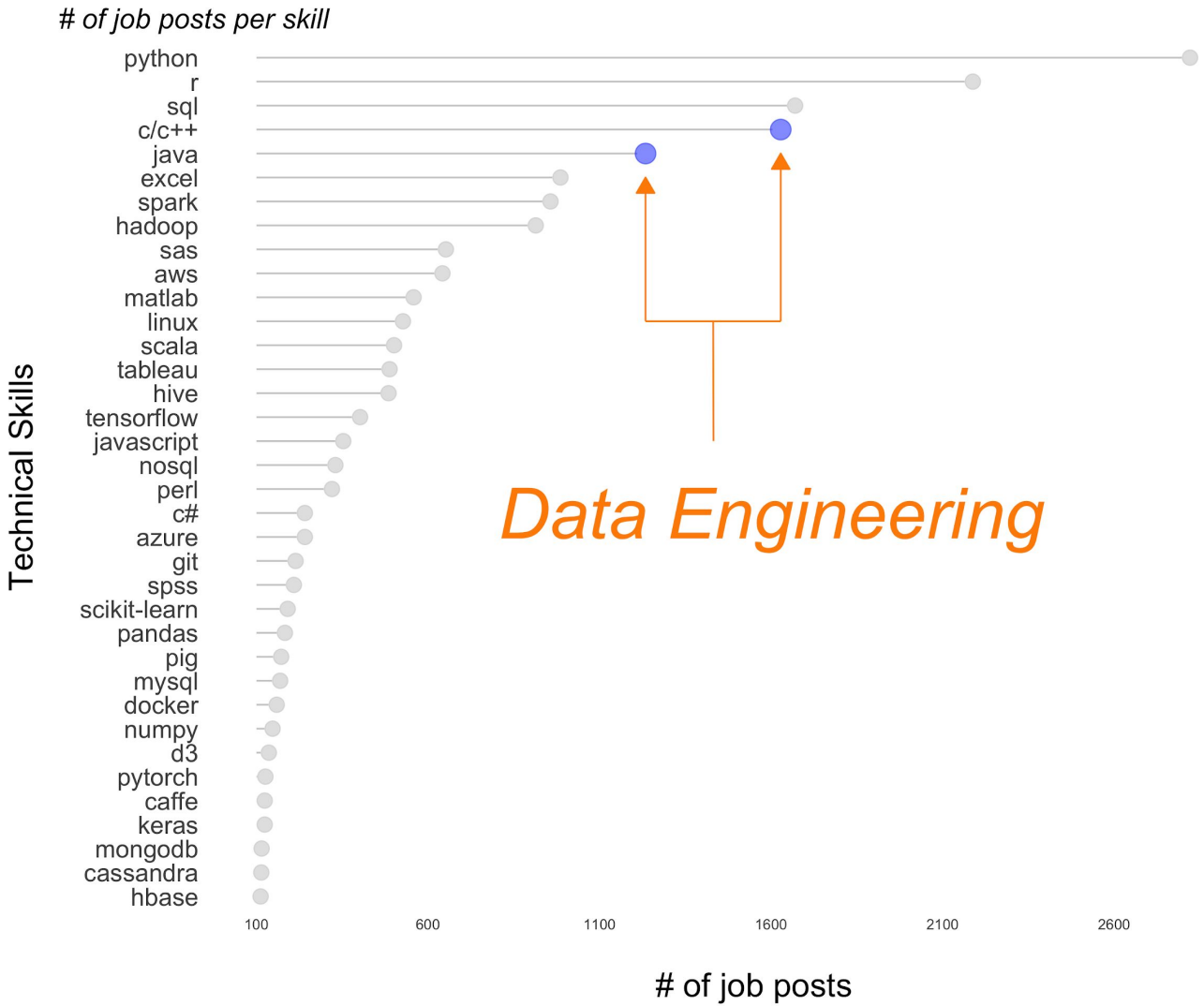

60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Job Percentage
Sponsor Jobs
Other
Data Scientist
Data Analyst
ML Engineer
Data Engineer

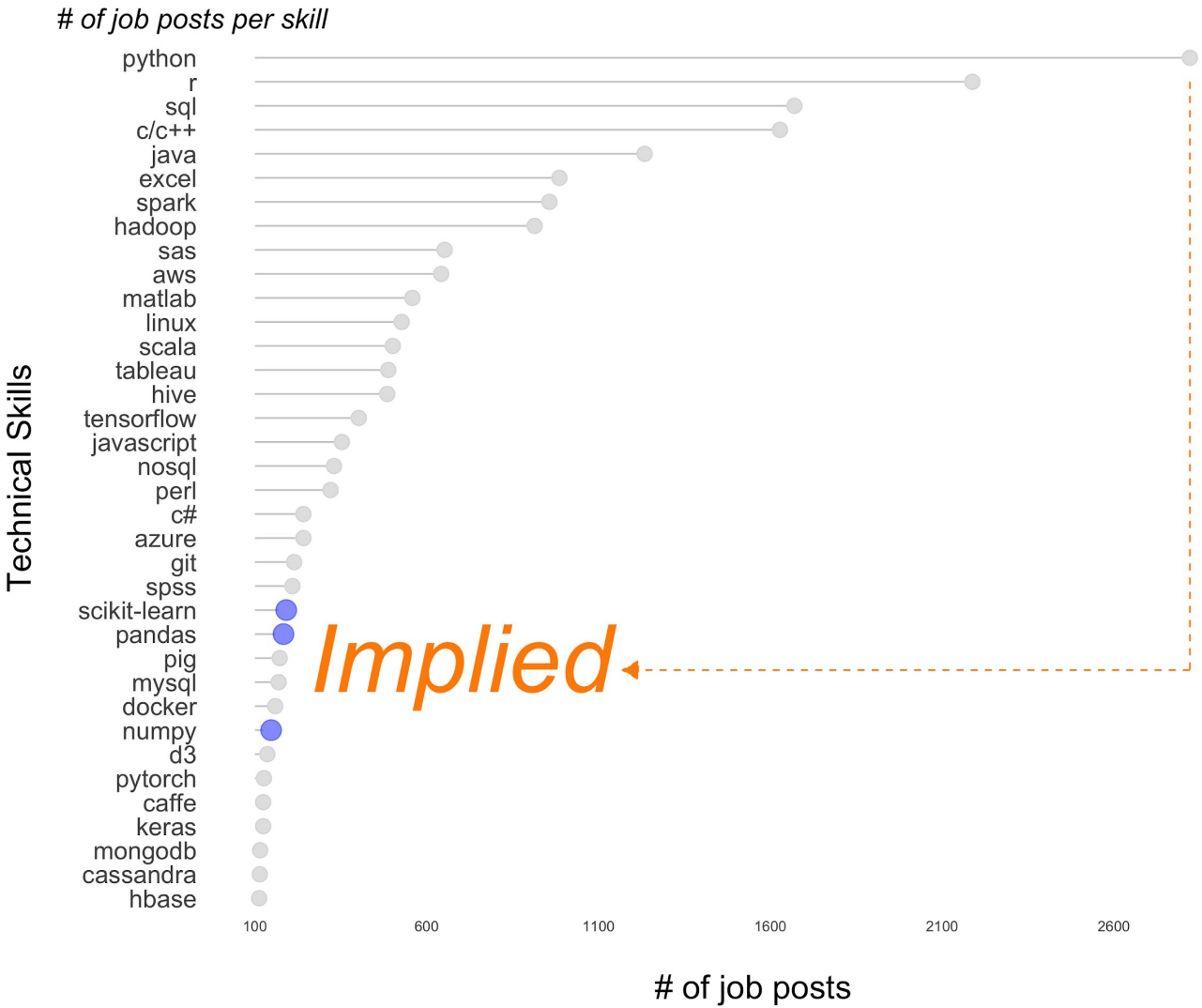
Number of DS Jobs per State

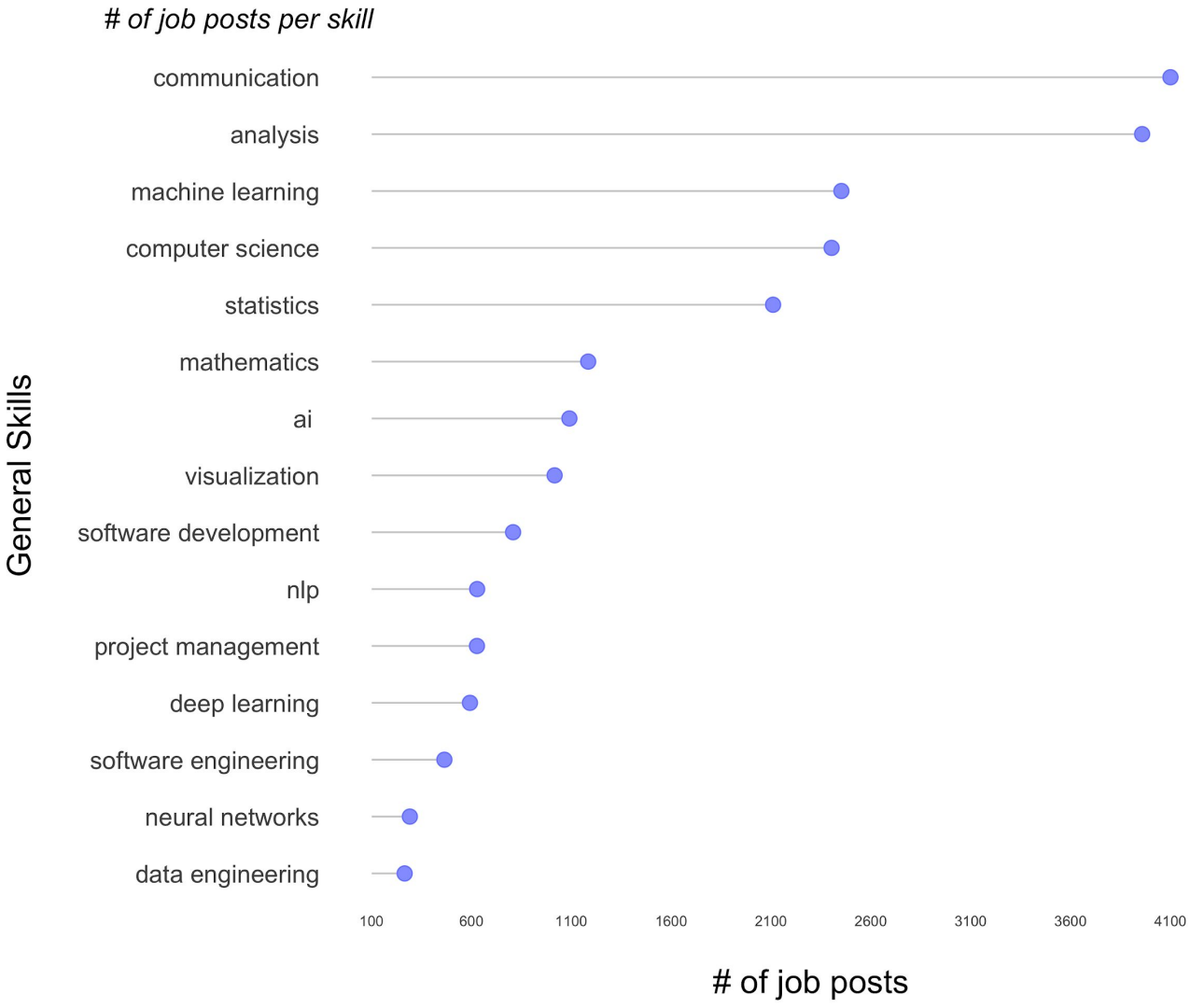


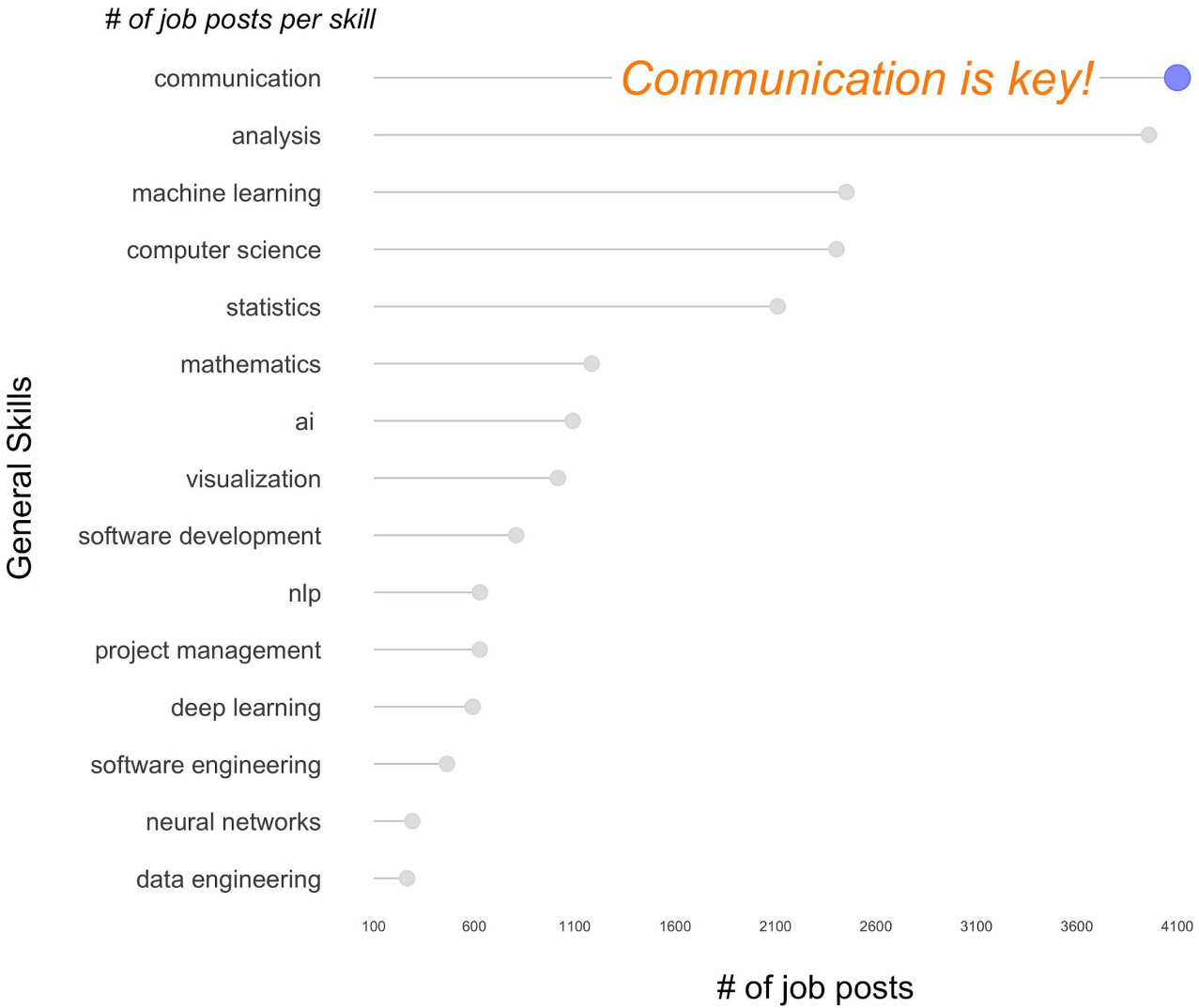
research
data
science
manager
development
business intelligence
engineering
machine
modeling
director
finance
project
program
systems
lead
reporting
chemistry
programming
ventures
automation
bioinformatics
biology
supply
chicago
department
execution
compensation
compliance
statistics
methodologist
coordinator
application
administration
product
information
consulting
media
safety
visualization
stability
solution
value
optimization
protein
credit
prevention
customer
cycle
payment
software
forecast
image
lab
tax
role
scene
team
cyber
logistics
records
level
mission
growth
decision
client
expert
measurement
biotech
revenue
support
delivery
healthcare
consultant
technology
performance
web
campaign
billing
improvement
control
capacity
trend
associate
market
specialist
risk
bioanalytics
process
equity
evaluation
geo
pricing
validation
president
people
network
developer
intelligence
operation
vice
population
design
strategy
policy
security
computer
therapy
crime
platform
sales
insight
quality
services
chain
trading
teaching
energy
formulation
water
consumer
investment
biochemistry
content
engagement
reimbursement
translation
benchmarks
knowledge
assistant
entry
linguist
multimodal
integrity
language
insurance
aerospace
initiatives
spectrometry
bureau
medicine
imaging
index
devops
enterprise
cybersecurity
periscope
infrastructure
climate
editor
impact
writing
phenotypic
facilities
method
welfare
bank
deals
property
vehicles
updates
tableau
wells
trust
search
discovery
physics
opportunity
genomics
bioassay
webspam
transaction
histology
head
study
derivatives
materials
functionality
drug
contract
fraud
tradeops
restaurants
planning
source
mile
immuno
threat
hazard
utility
identity
mining
valuations
tier
secret
trader
access
time
geo
inventory
lending
biomarker
orders
clearance
institute
pharmacovigilance
gaming
partnerships
foundation
mychart
tradecraft
biologics
journey
estate
chromatography
professor
outsourcing
issue
asset
ecommerce
life sciences
flight
stack
child
repertoire
defense
personalization
center
experiments
graduate
manger
oncology
models
recruiter
career
malware
capital
mass
morphometry
outcomes
underwriting
advertising
locations
migration
securities
ignition
graph
treasury
accounting
database
liaison
review
income
ops
warehouse
treatment
claims
python
signal
officer
fellow
subways
custom
preclinical
sls
cancer
vaccines
cloud
video
text
success
school
procurement
yield
sensing
assessment
diversity
volatility
conversion
instructor
understanding
options
training
mapping
frequency
course
infrastructure
periscope
editor
cybersecurity
devops
enterprise
imaging
index
medicine
spectrometry
bureau
initiatives
aerospace
insurance
integrity
language
multimodal
linguist
engagement
reimbursement
entry
assistant
translation
benchmarks
knowledge
impact
writing
phenotypic
facilities
method
search
discovery
physics
opportunity
genomics
bioassay
webspam
transaction
histology
head
study
derivatives
materials
functionality
drug
contract
fraud
tradeops
restaurants
planning
source
mile
immuno
threat
hazard
utility
identity
mining
valuations
tier
secret
trader
access
time
geo
inventory
lending
biomarker
orders
clearance
institute
pharmacovigilance
gaming
partnerships
foundation
mychart
tradecraft
biologics
journey
estate
chromatography
professor
outsourcing
issue
asset
ecommerce
life sciences
flight
stack
child
repertoire
defense
personalization
center
experiments
graduate
manger
oncology
models
recruiter
career
malware
capital
mass
morphometry
outcomes
underwriting
advertising
locations
migration
securities
ignition
graph
treasury
accounting
database
liaison
review
income
ops
warehouse
treatment
claims
python
signal
officer
fellow
subways
custom
preclinical
sls
cancer
vaccines
cloud
video
text
success
school
procurement
yield
sensing
assessment
diversity
volatility
conversion
instructor
understanding
options
training
mapping
frequency
course
infrastructure
periscope
editor
cybersecurity
devops
enterprise
imaging
index
medicine
spectrometry
bureau
initiatives
aerospace
insurance
integrity
language
multimodal
linguist
engagement
reimbursement
entry
assistant
translation
benchmarks
knowledge
impact
writing
phenotypic
facilities
method
search
discovery
physics
opportunity
genomics
bioassay
webspam
transaction
histology
head
study
derivatives
materials
functionality
drug
contract
fraud
tradeops
restaurants
planning
source
mile
immuno
threat
hazard
utility
identity
mining
valuations
tier
secret
trader
access
time
geo

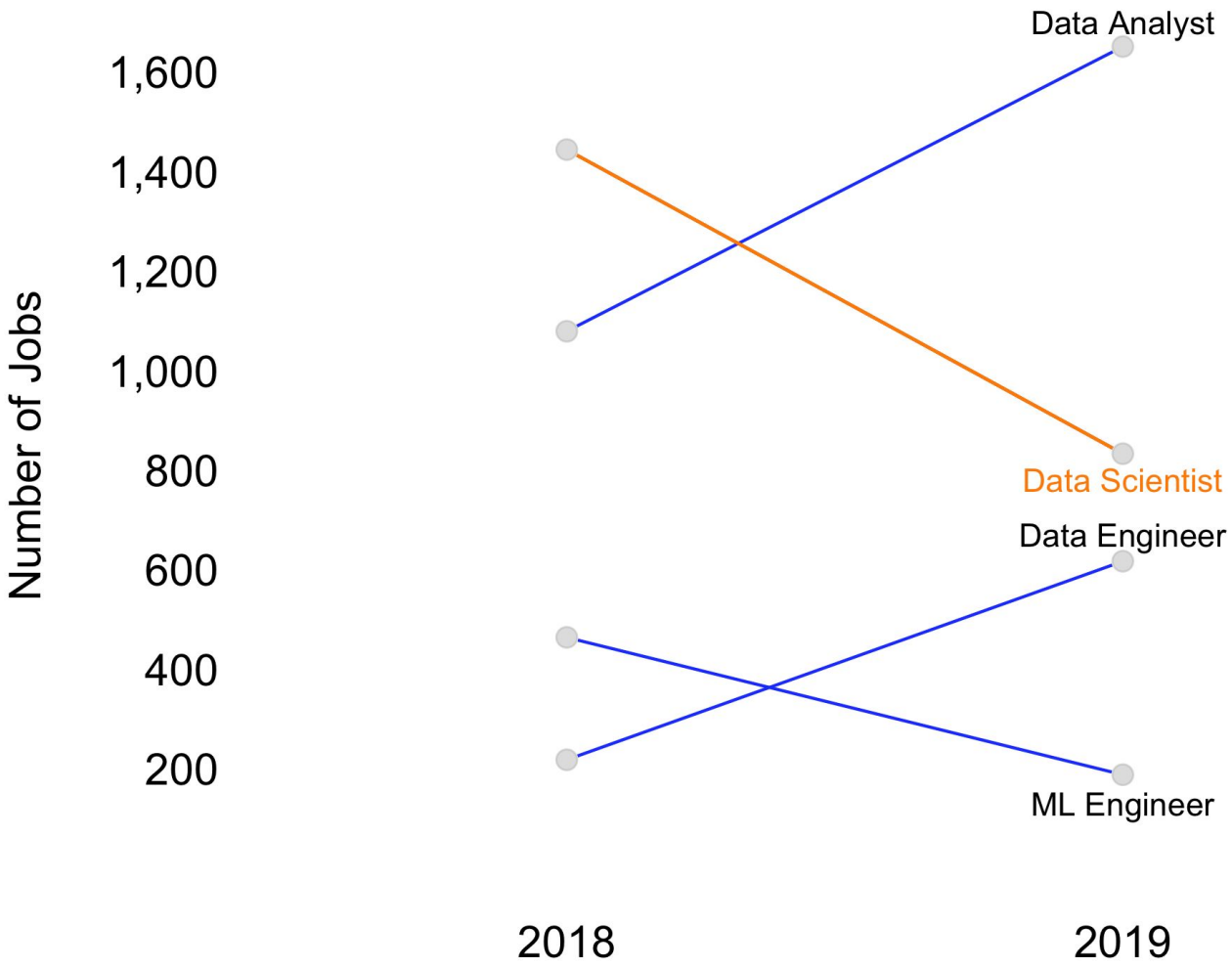
# of job posts per skill
python
r
sql
c/c++
java
excel
spark
hadoop
sas
aws
matlab
linux
scala
tableau
hive
tensorflow
javascript
nosql
perl
c#
azure
git
spss
scikit-learn
pandas
pig
mysql
docker
numpy
d3
pytorch
caffe
keras
mongodb
cassandra
hbase
Technical Skills
100
600
1100
1600
2100
2600
# of job posts
Learning many of these right now!


# of job posts per skill
python
r
sql
c/c++
java
excel
spark
hadoop
sas
aws
matlab
linux
scala
tableau
hive
tensorflow
javascript
nosql
perl
c#
azure
git
spss
scikit-learn
pandas
pig
mysql
docker
numpy
d3
pytorch
caffe
keras
mongodb
cassandra
hbase
Technical Skills
100
600
1100
1600
2100
2600
# of job posts
Data Engineering


# of job posts per skill
Technical Skills
python
r
sql
c/c++
java
excel
spark
hadoop
sas
aws
matlab
linux
scala
tableau
hive
tensorflow
javascript
nosql
perl
c#
azure
git
spss
scikit-learn
pandas
pig
mysql
docker
numpy
d3
pytorch
caffe
keras
mongodb
cassandra
hbase
Implied
100
600
1100
1600
2100
2600
# of job posts


# of job posts per skill
communication
analysis
machine learning
computer science
statistics
mathematics
ai
visualization
software development
nlp
project management
deep learning
software engineering
neural networks
data engineering
General Skills
100
600
1100
1600
2100
2600
3100
3600
4100
# of job posts


# of job posts per skill
Communication is key!
communication
analysis
machine learning
computer science
statistics
mathematics
ai
visualization
software development
nlp
project management
deep learning
software engineering
neural networks
data engineering
General Skills
100
600
1100
1600
2100
2600
3100
3600
4100
# of job posts


Number of Jobs
1,600
1,400
1,200
1,000
800
600
400
200
2018
2019
Data Analyst
Data Scientist
Data Engineer
ML Engineer

Summary

- Majority of Data Science related roles are located in CA > NY > MA > WA
- Don’t be afraid to look into Data Analyst and Data Engineer
- Most important technical skills: Python, SQL
- Most important general skill: Communication

Thank you!