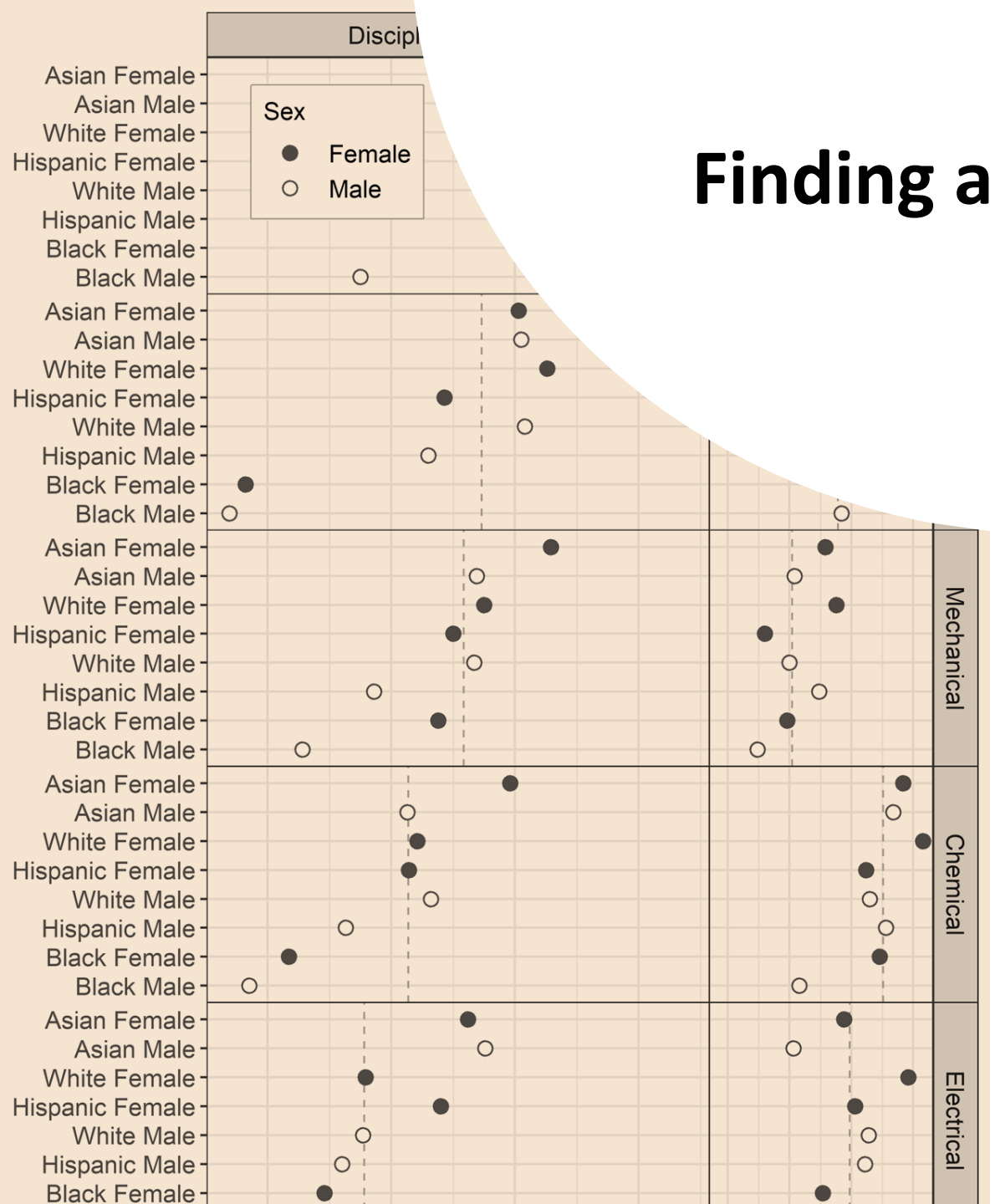


Finding and presenting stories in the data

2019 MIDFIELD Institute

Susan Lord
Richard Layton

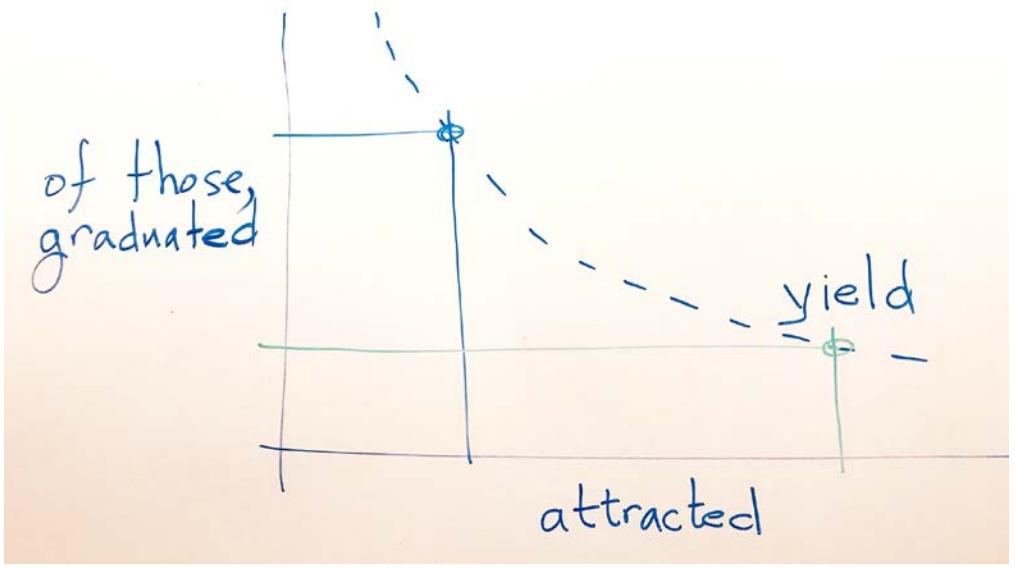
Tuesday, June 4, 2019

MIDFIELD
Multiple-Institution Database For Investigating Engineering Longitudinal Development

This talk focuses on finding the stories in the data

version 1
version 2
version 3

Design of effective displays is iterative



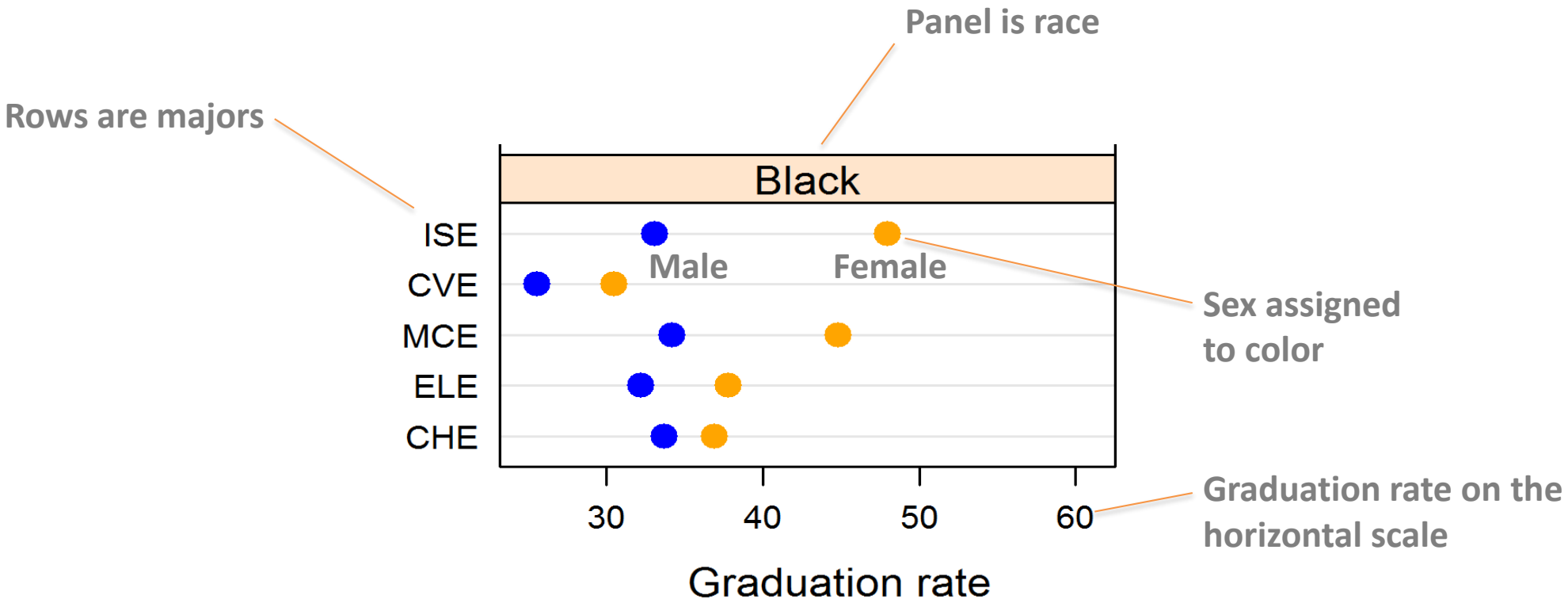
Complex displays start with simple conceptual sketches

Example 1

Iteratively exploring the starters' graduation rates

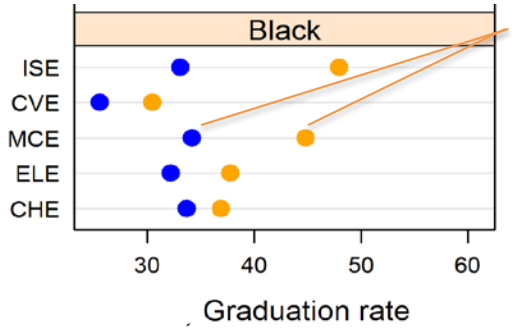
In an initial look at graduation rate, we placed men and women on the same row.

version 1



A multiway design facilitates comparisons within a panel.

version 1

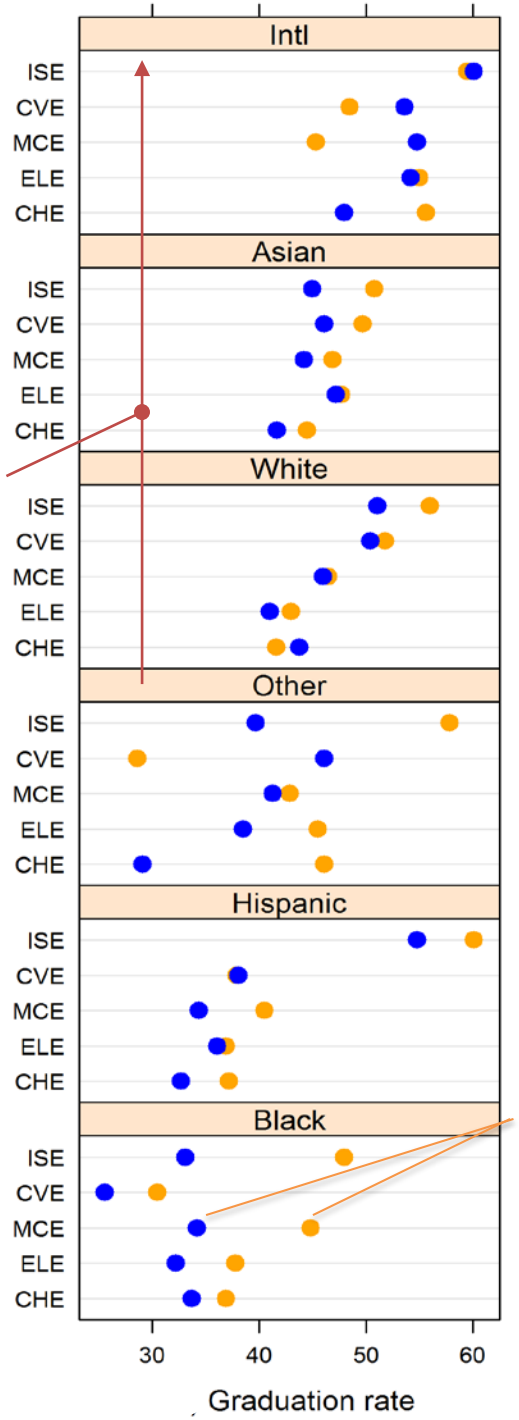


Compare by **sex**

One **race**,
compare by **major**

A multiway design facilitates comparisons within a panel.

version 1



Order panels by median **by race**

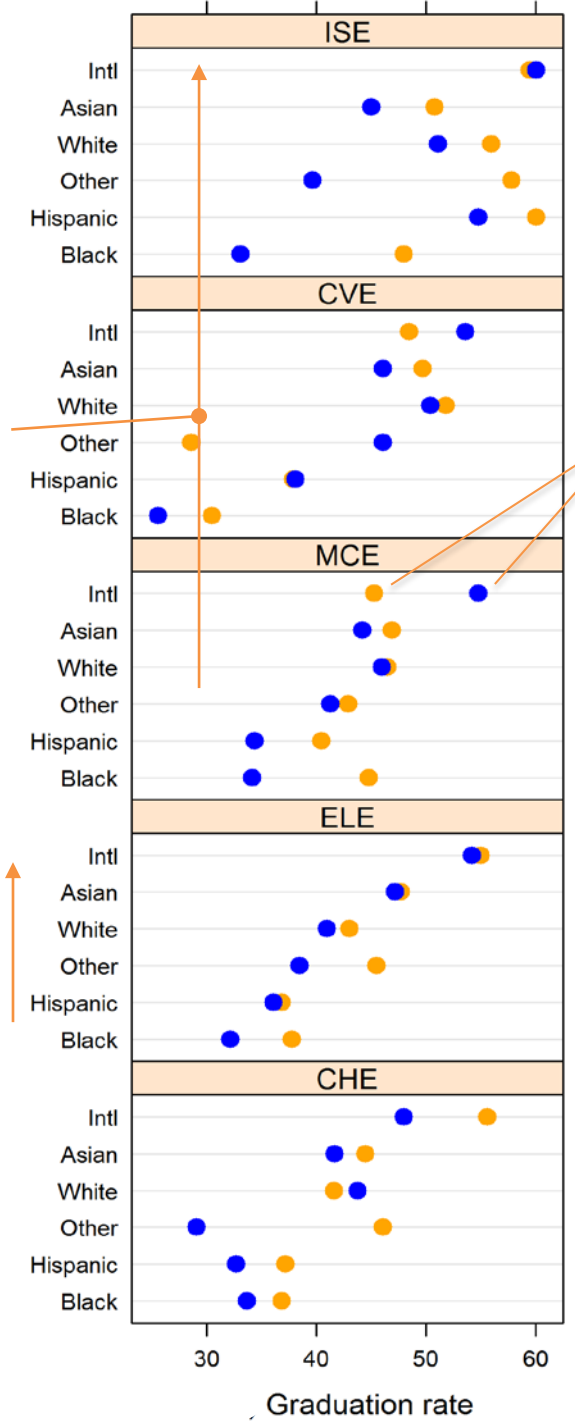
Order rows by median **by major**

Compare by **sex**

One **race**, compare by **major**

In the dual multiway, we swap the roles of rows and panels.

version 2



Order panels by median **by major**

Order rows by median **by race**

Compare by **sex**

One **major**, compare **by race**

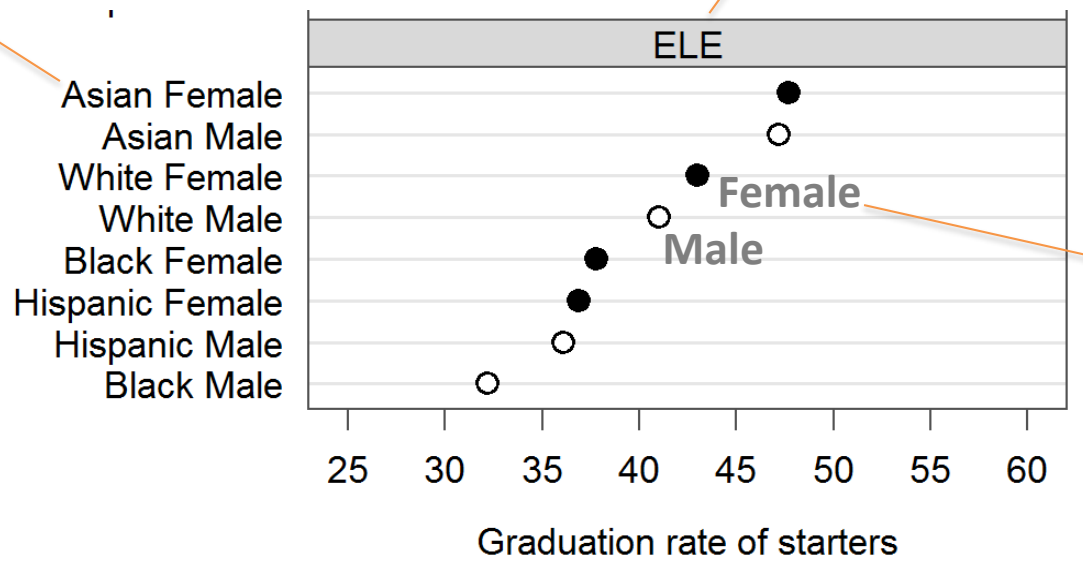
The next iteration placed men and women on separate rows

version 3

Rows are race + sex

We omit International and Other

Panel is major

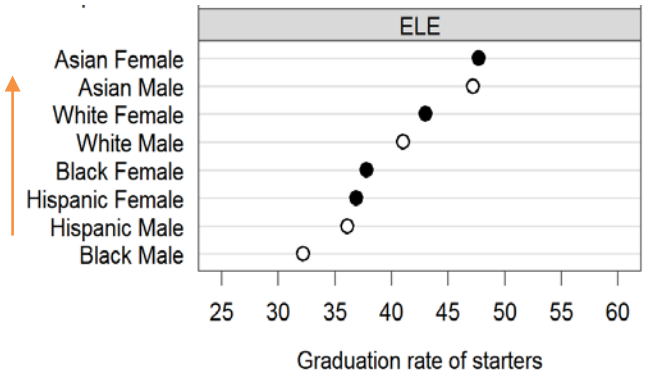


Sex assigned to symbol type

In this design, the rows are ordered by the data in the panel

version 3

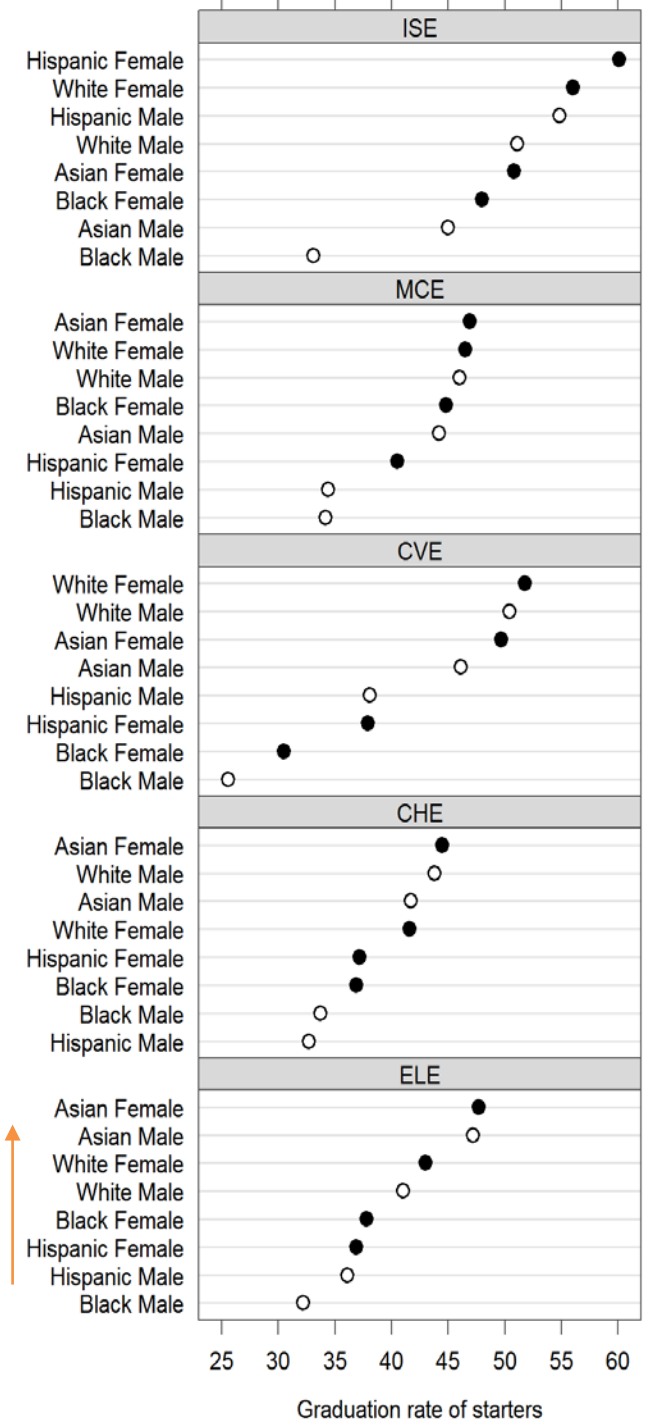
Order rows by data in panel



One major, compare by race/sex

In this design, the rows are ordered by the data in the panel

version 3



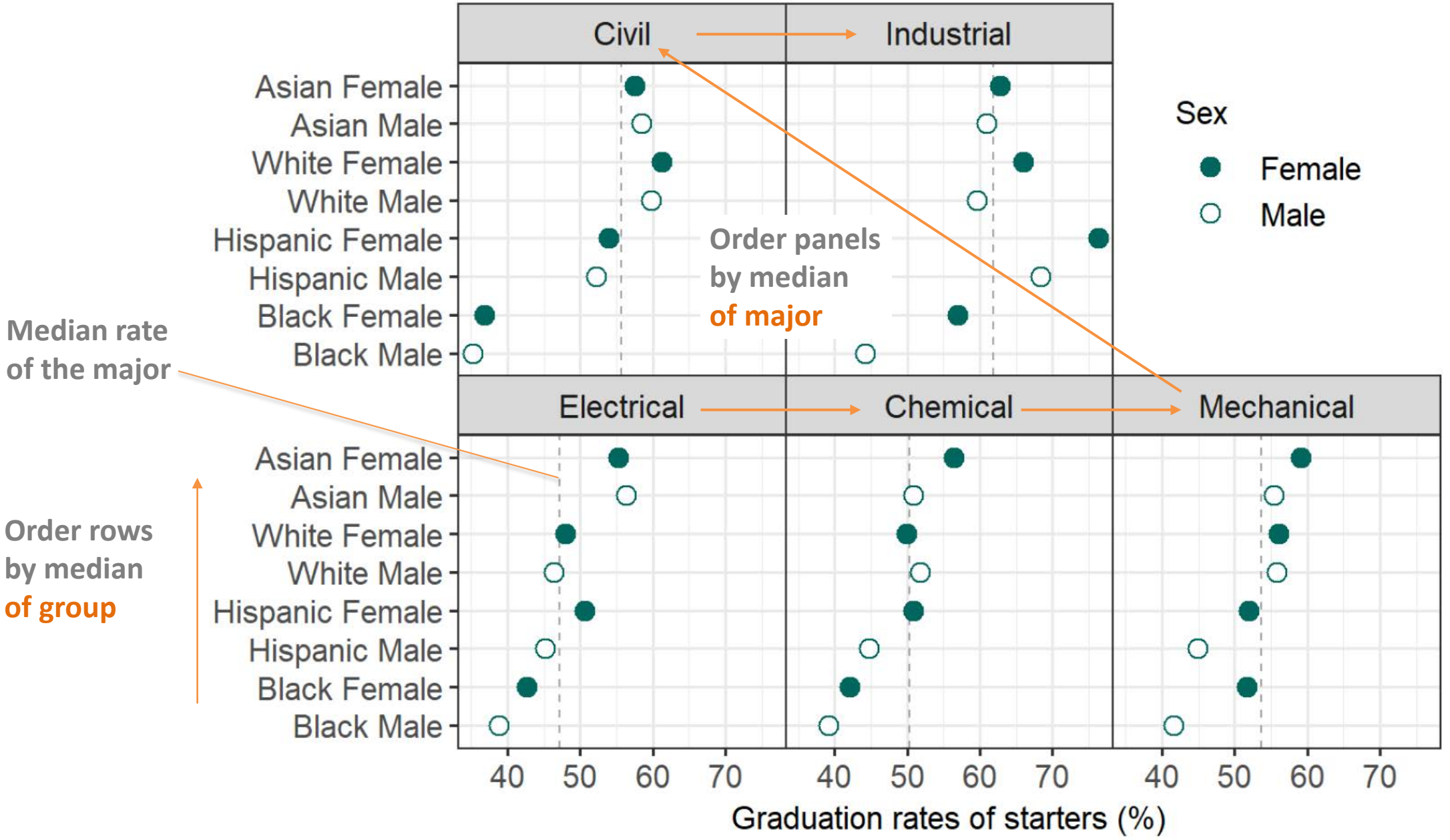
Order panels by median **in the panel**

Order rows by data **in panel**

One **major**, compare **by race/sex**

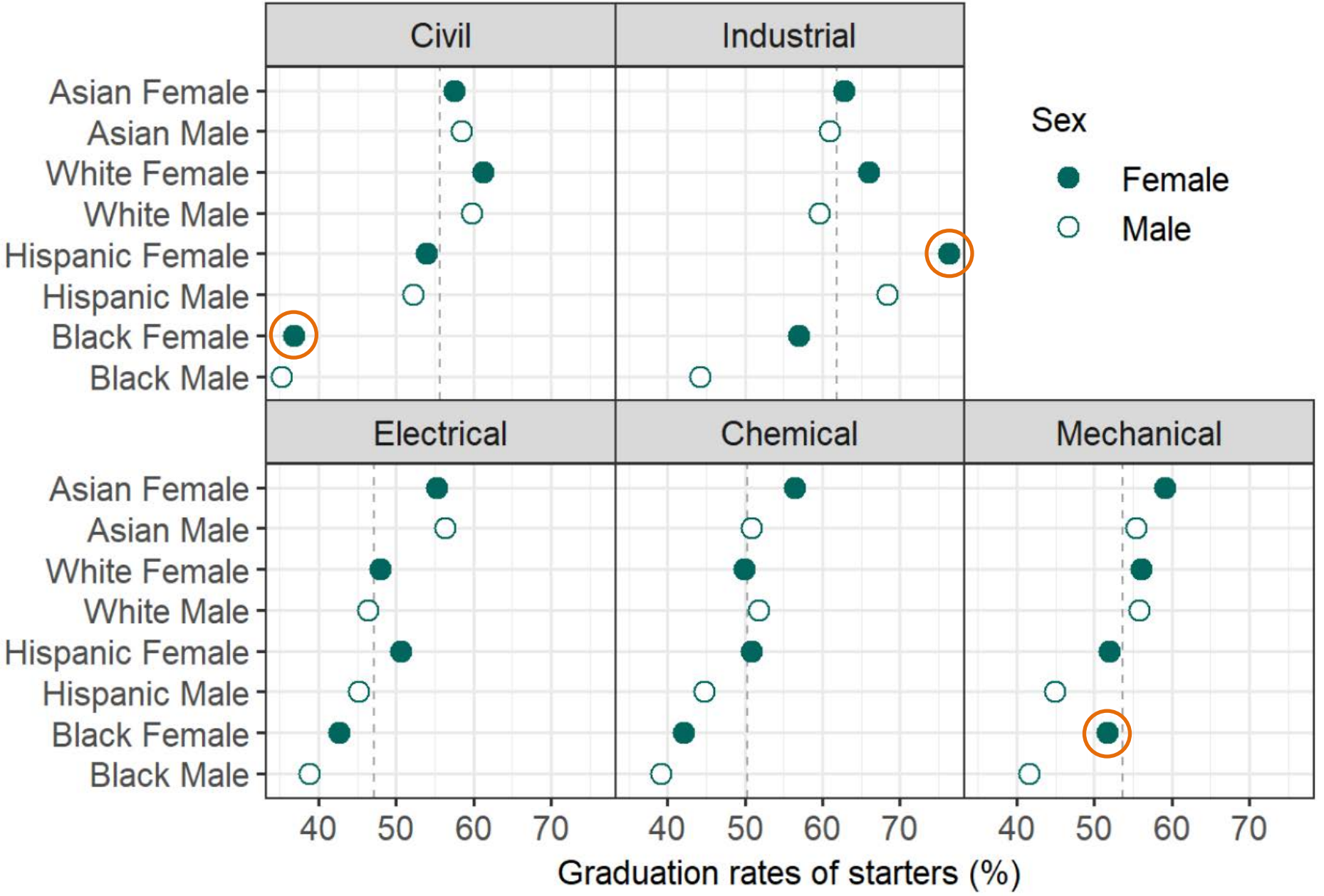
In the final design, all rows are in the same order

version 4



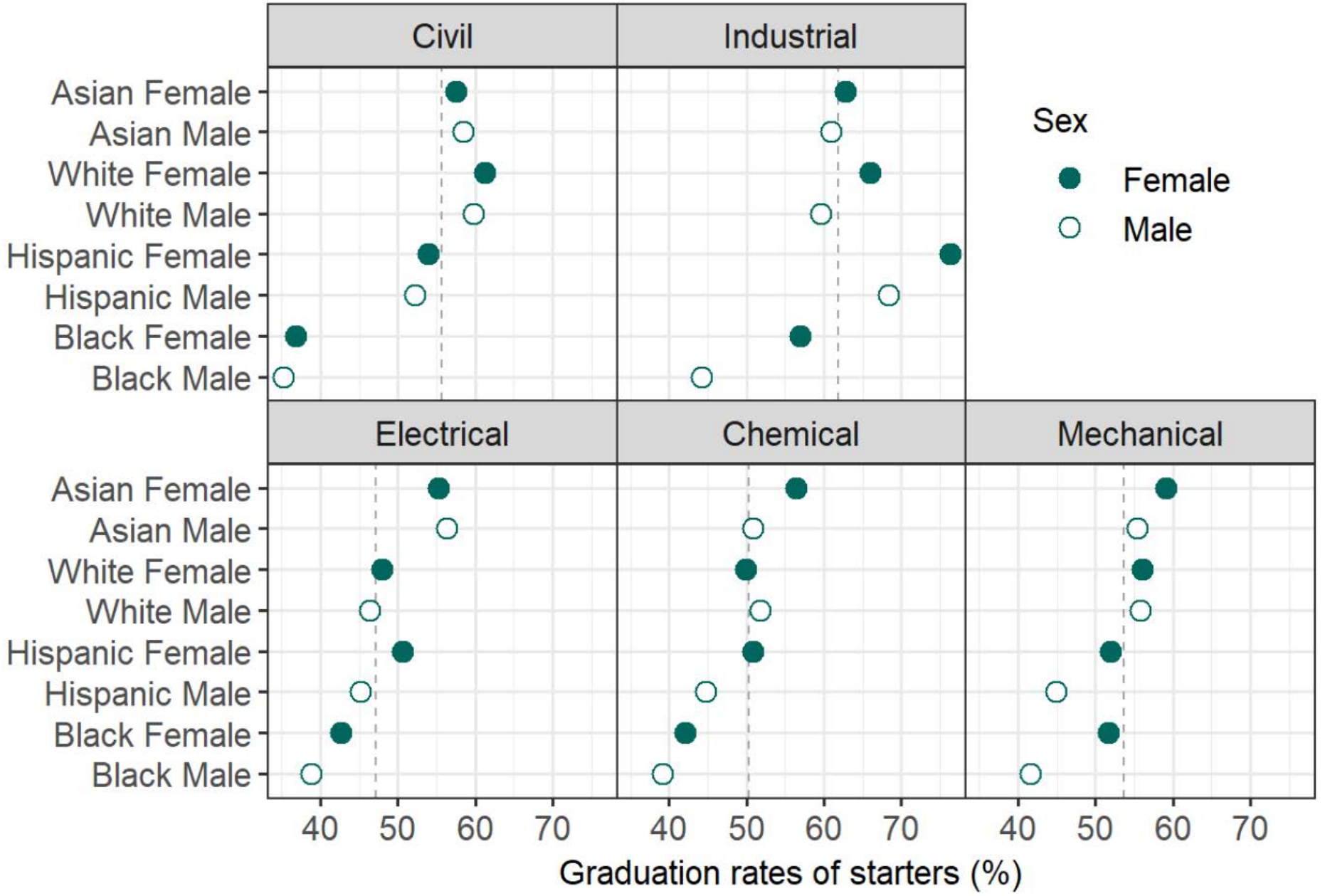
This multiway design highlights visual anomalies.

version 4



This multiway design highlights visual anomalies.

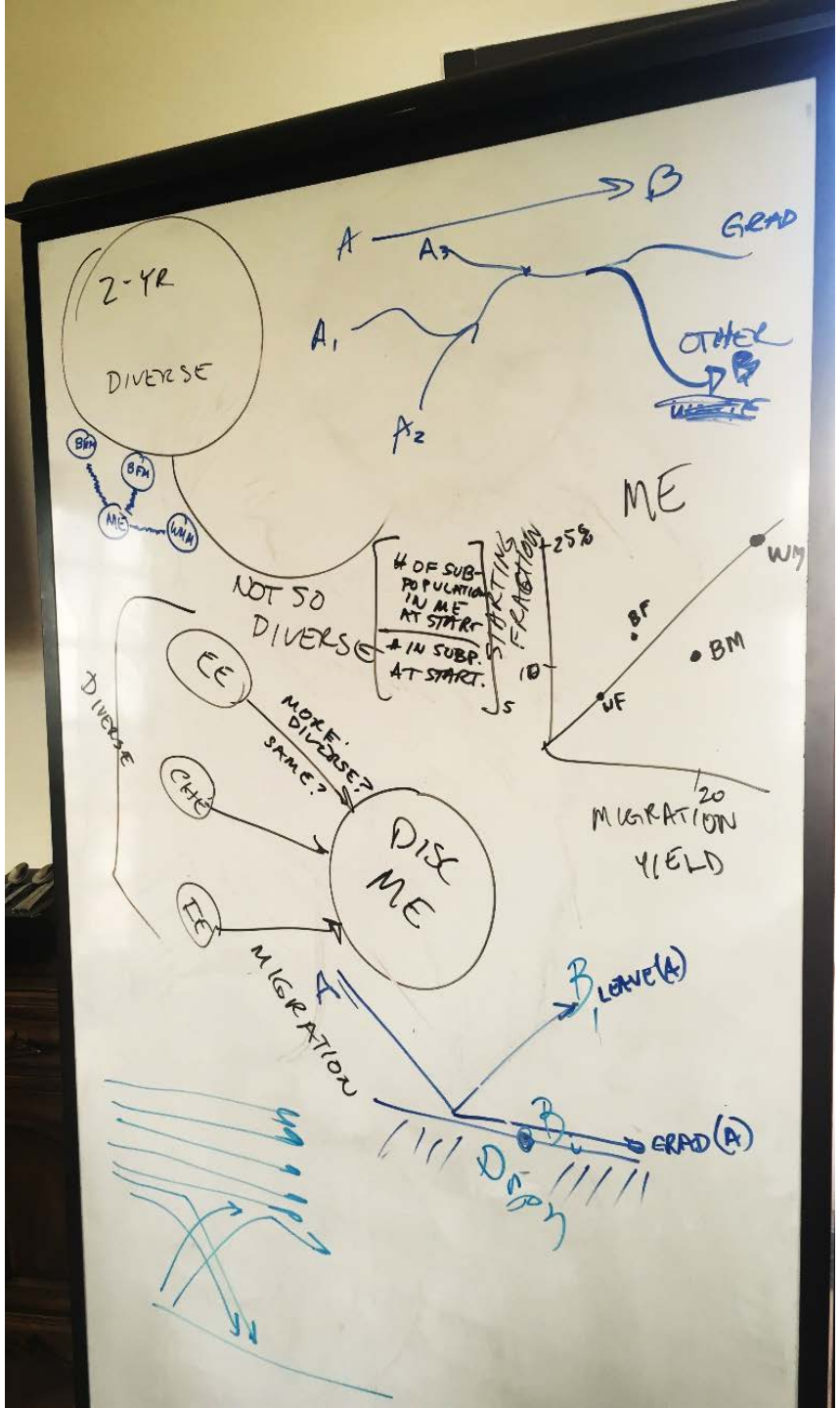
version 4



Example 2

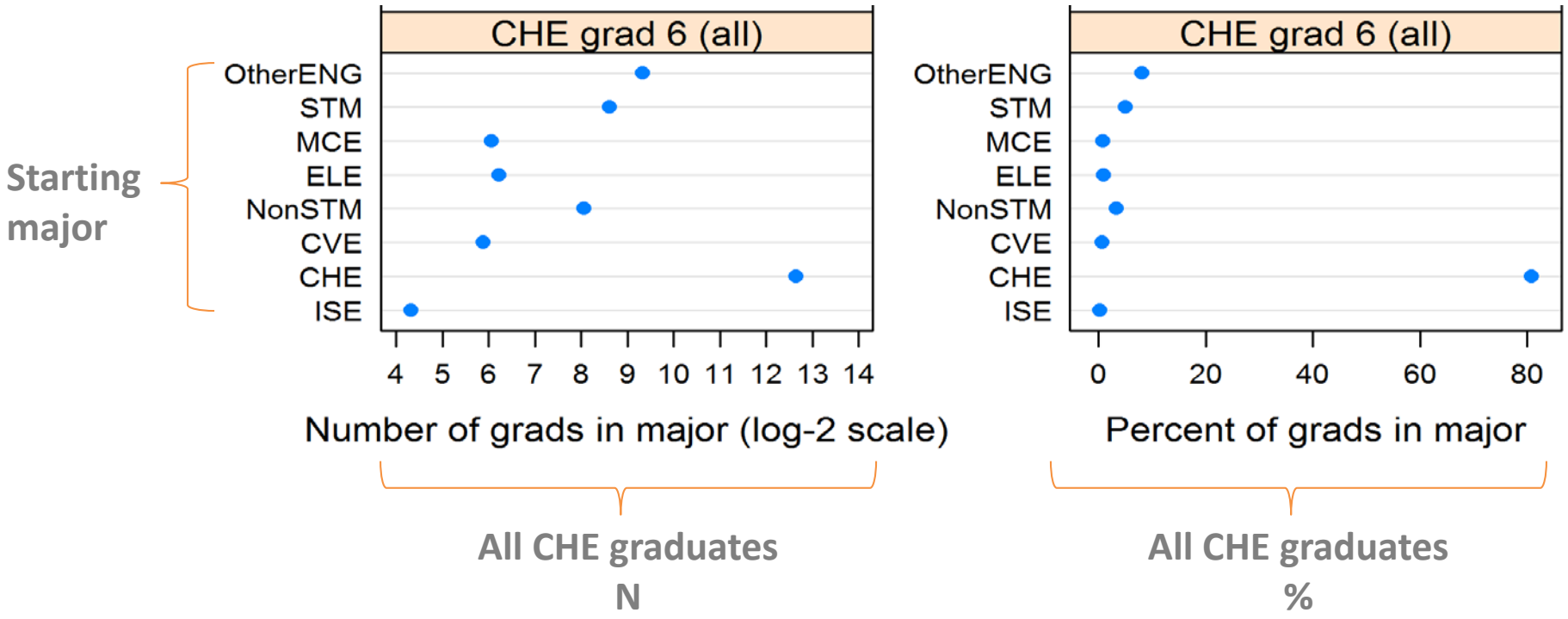
Iteratively exploring the migrators' stories

Initially we sketched and discussed around a white board.



In our first attempt we asked where grads in a major started

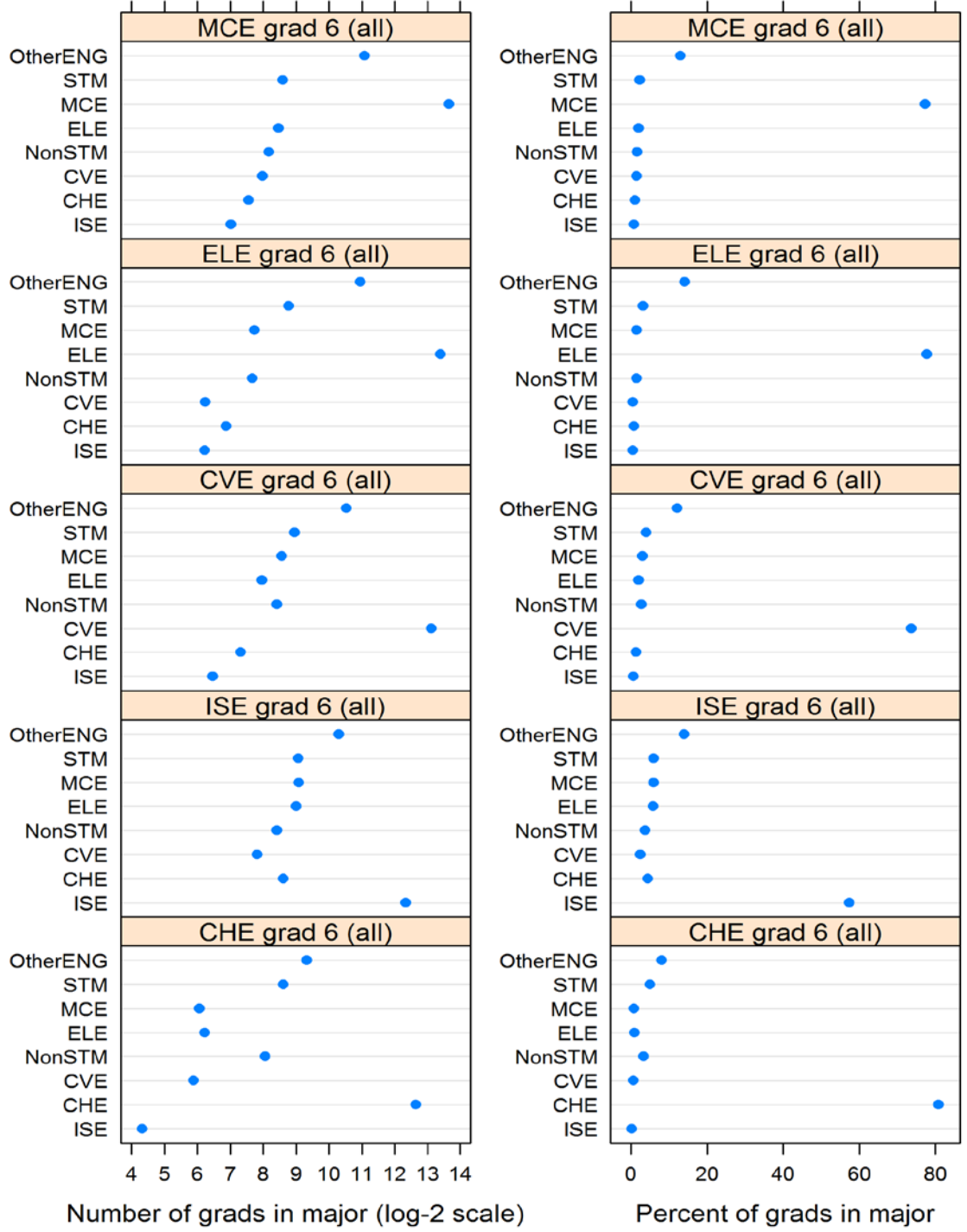
version 1



Same graph, 5 majors

version 1

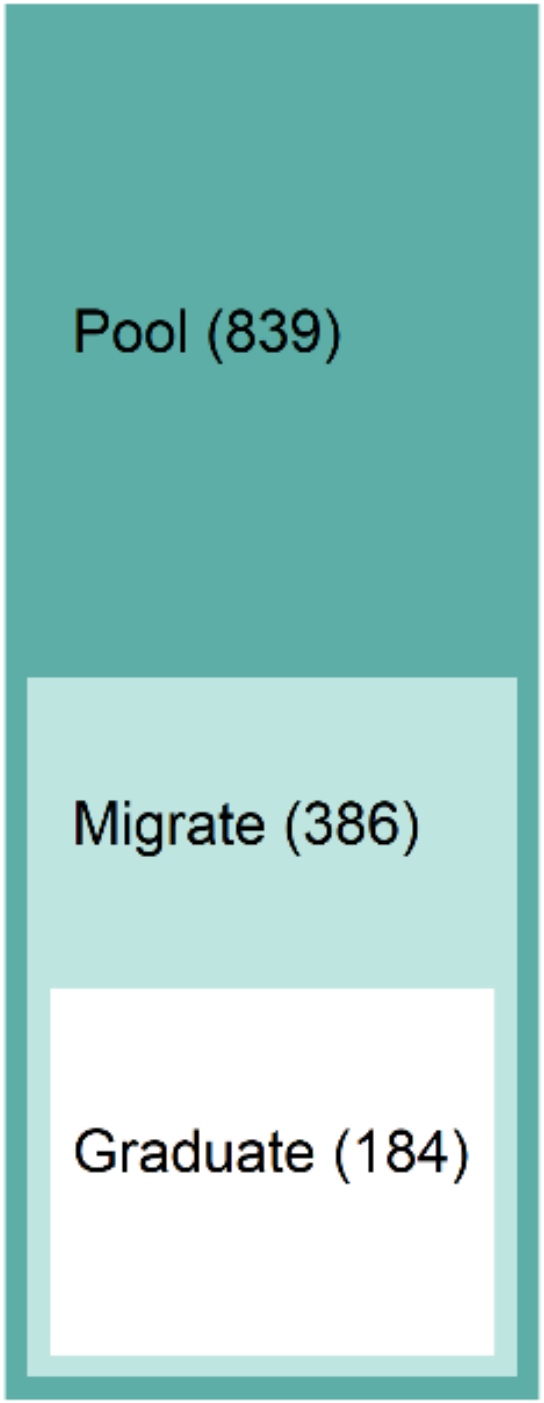
The visual story didn't seem meaningful



We constructed a new metric

A new metric, migration yield, was hinted at in our initial brainstorming.

We constructed a new metric



Black Male students in EE

839 potential migrators to EE

Of those, 386 migrated to EE

Fraction of migrators attracted = 386 / 839 = 0.460

Of those, 184 graduated in EE

Fraction attracted that graduate = 184 / 386 = 0.477

Migration yield is computed in one of two ways:

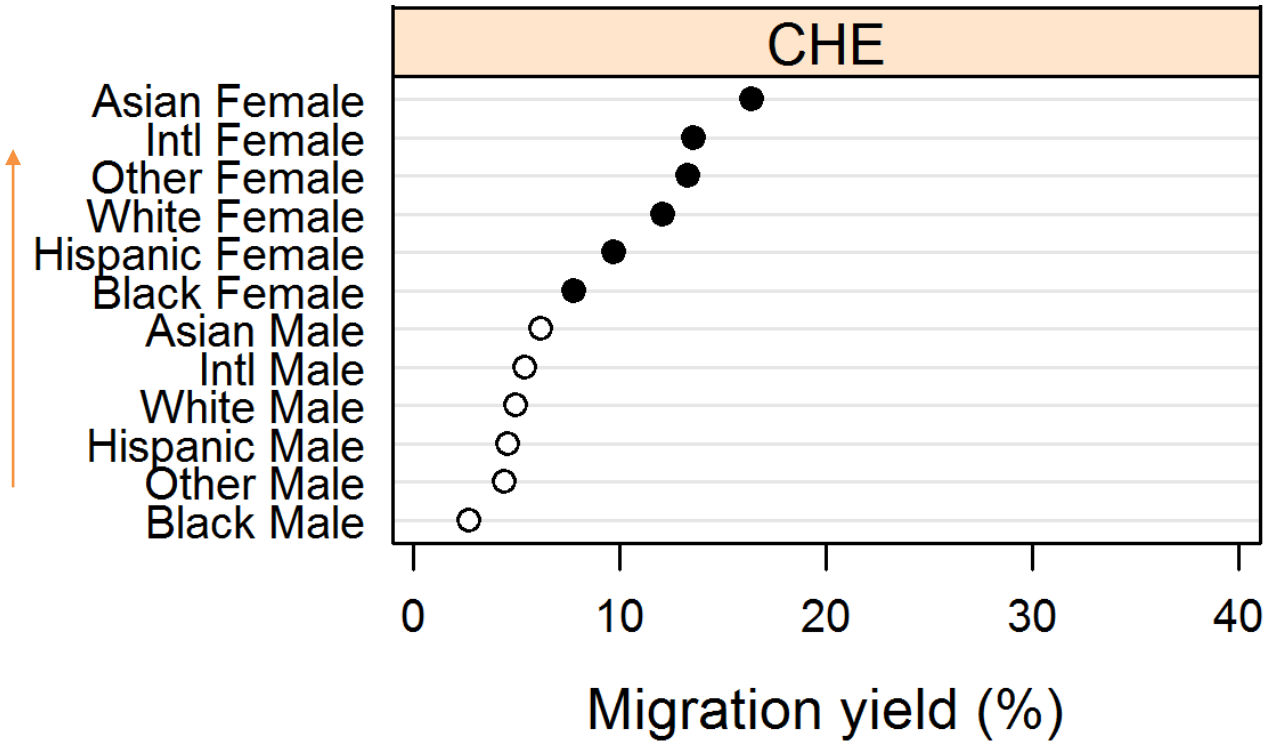
Product of the two fractions: 0.460 x 0.477 = 22%

Ratio of graduates to pool: 184 / 839 = 22%

In the first attempt, we graphed migration yield directly

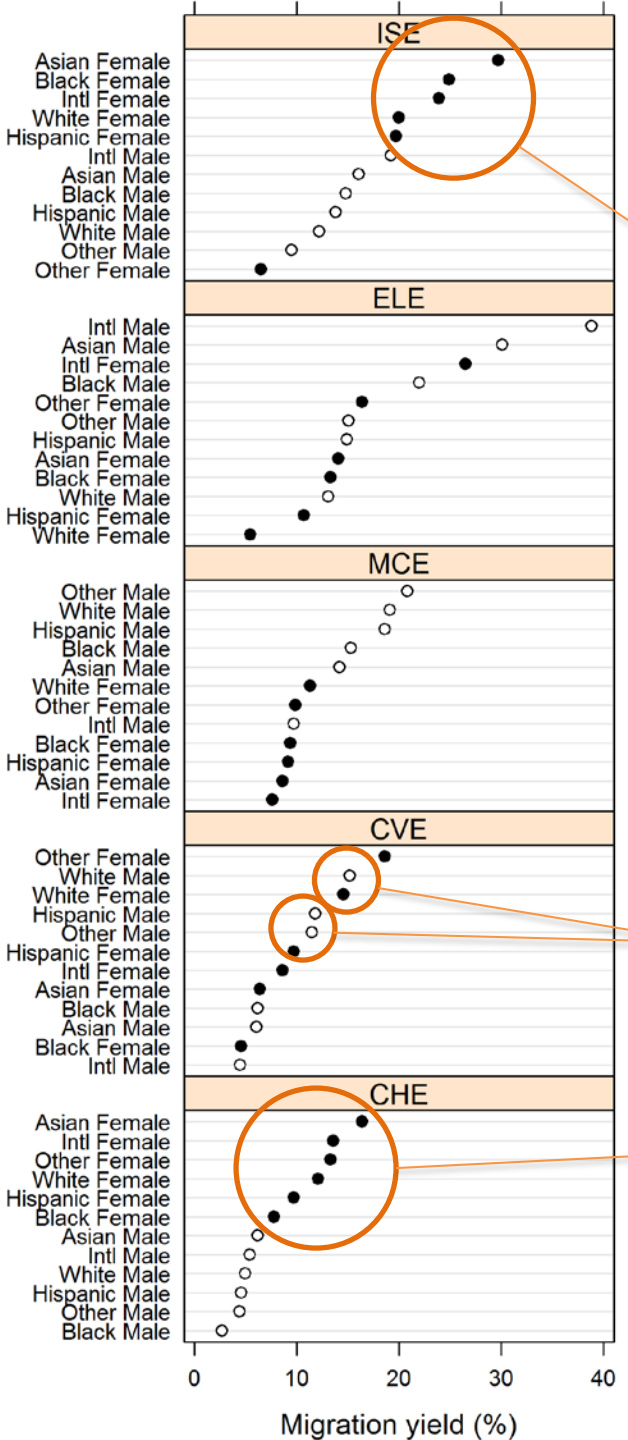
version 2

Order rows by the **data values**



Same graph, 5 majors

version 2



Clustered by **sex**

Order rows by the **data values**

Clustered by **race**

Clustered by **sex**

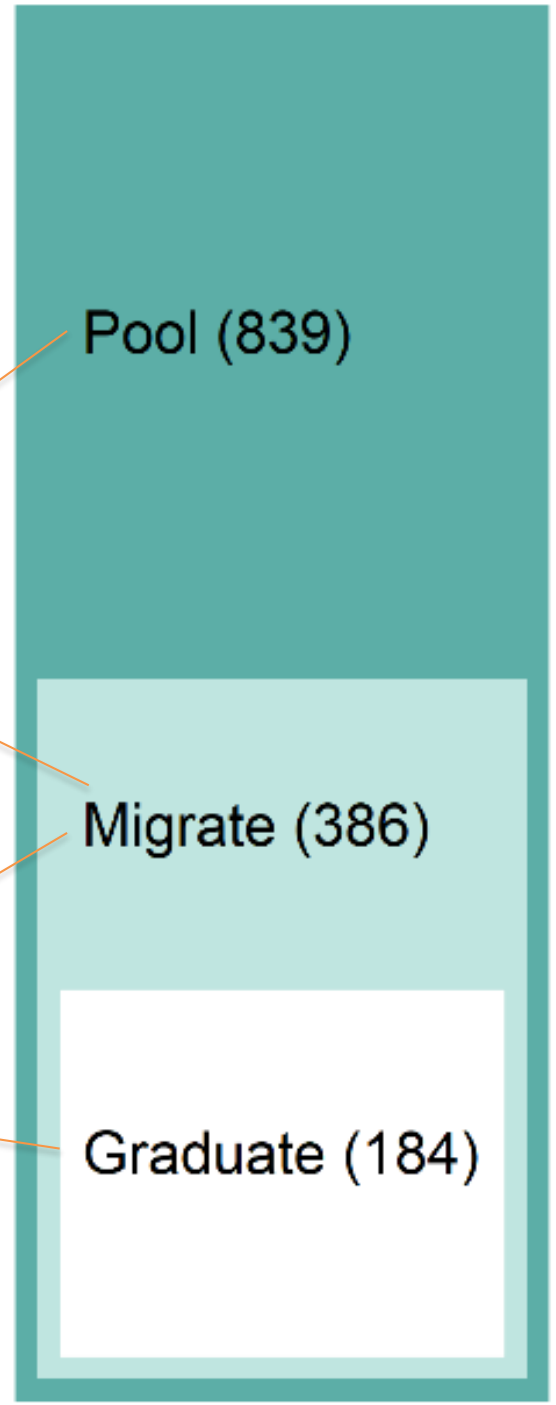
We realized we had two distinct concepts contributing to migration yield

The fraction of the pool who were attracted

×

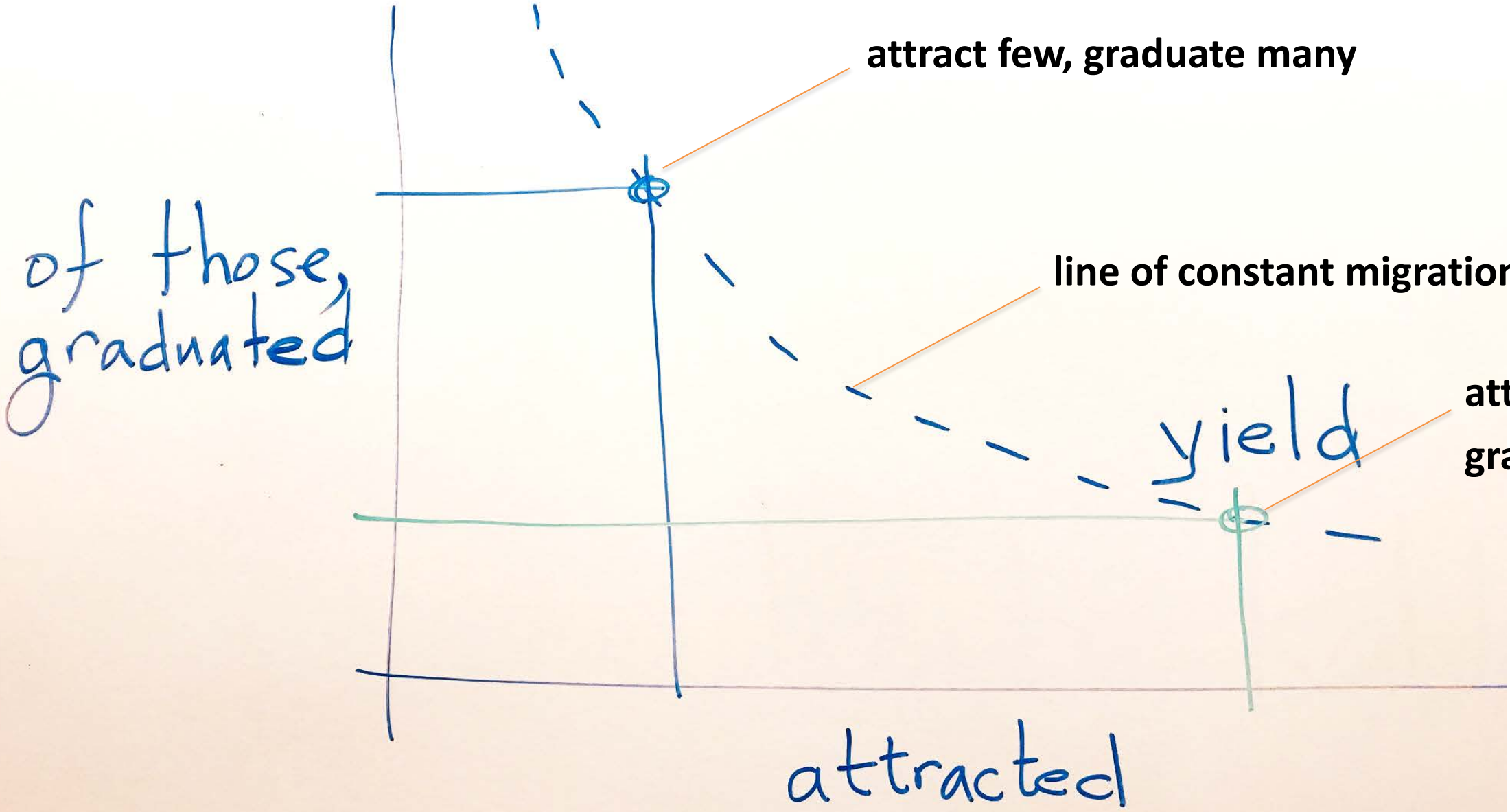
The fraction of those who graduated

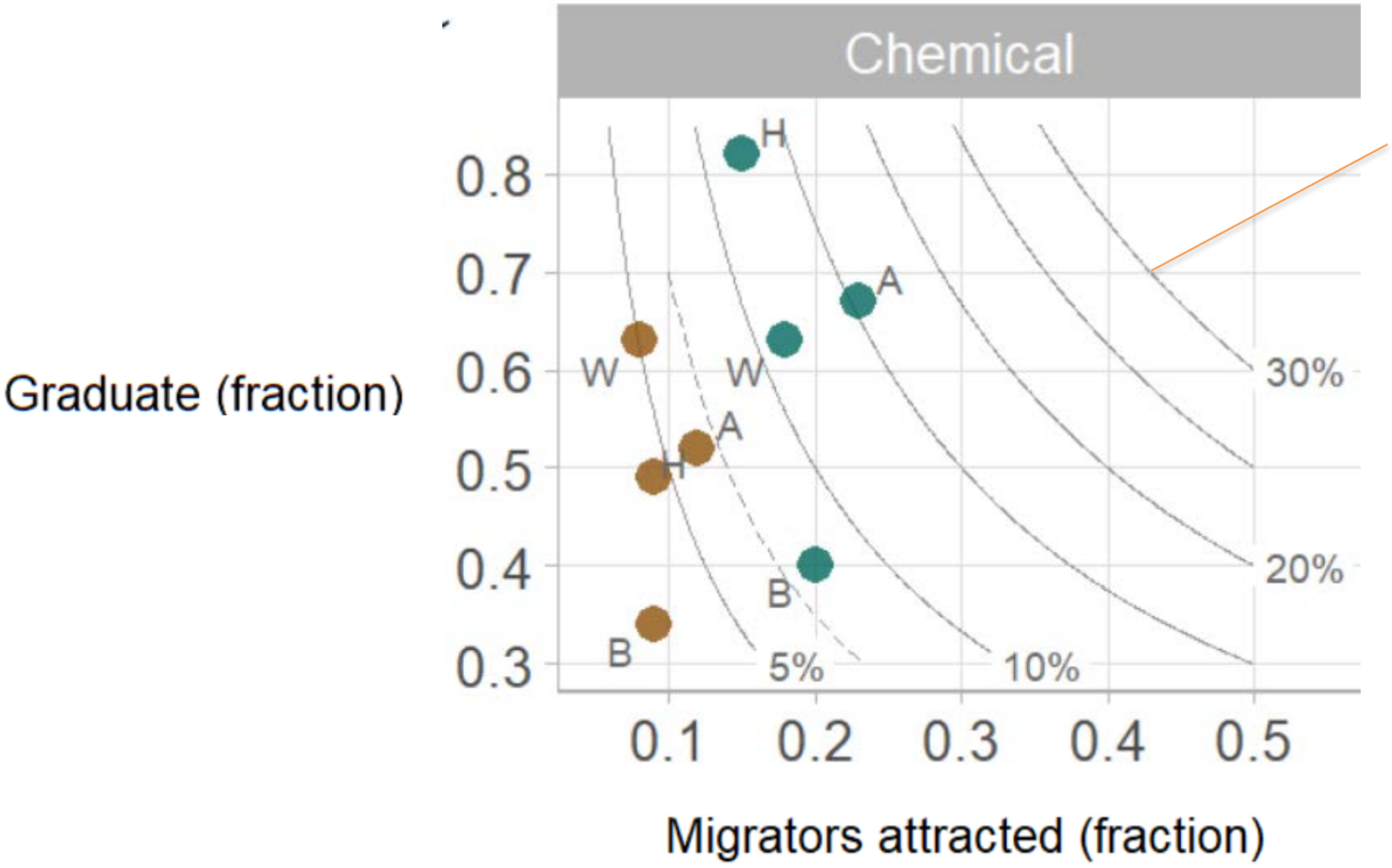
= migration yield



In a Cartesian graph, a constant product is a contour

version 3



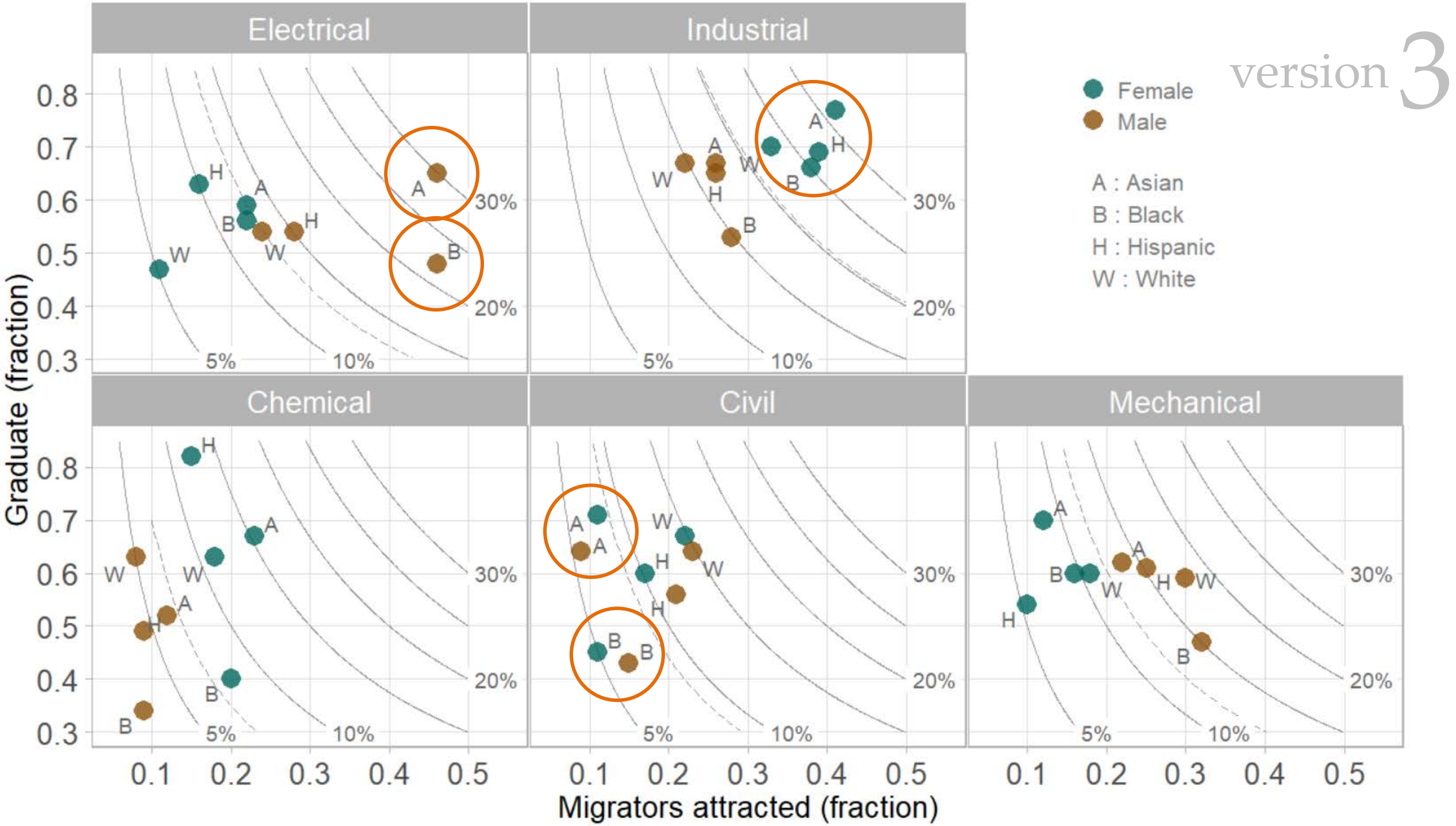


lines of constant migration yield

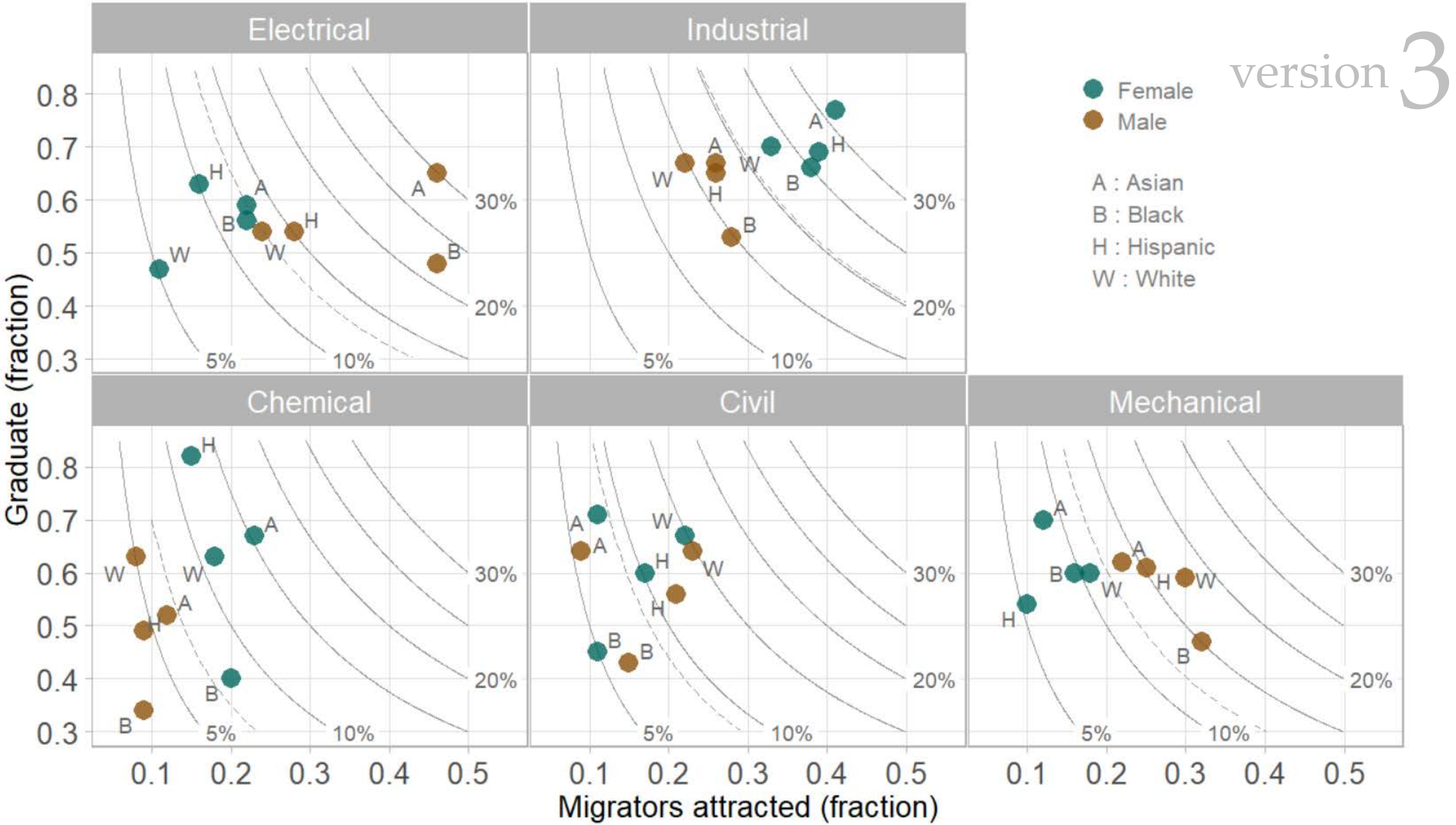
Female
Male

A : Asian
B : Black
H : Hispanic
W : White

two legends required

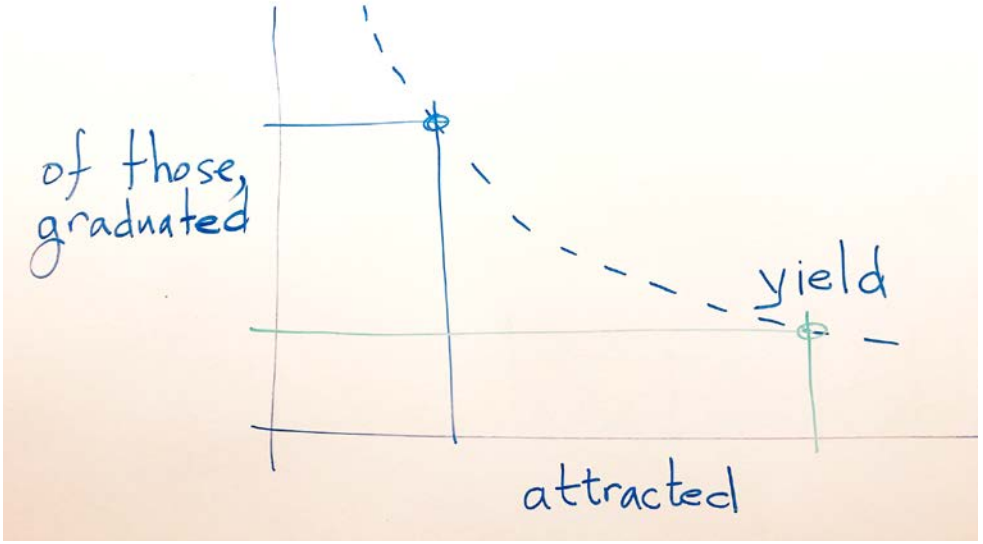

Electrical
Industrial
Chemical
Civil
Mechanical
Female
Male
A : Asian
B : Black
H : Hispanic
W : White
30%
20%
5%
10%
Graduate (fraction)
Migrators attracted (fraction)
0.3
0.4
0.5
0.6
0.7
0.8
0.1
0.2
0.3
0.4
0.5

version 3


version 3
Electrical
Industrial
Chemical
Civil
Mechanical
Female
Male
A : Asian
B : Black
H : Hispanic
W : White
Graduate (fraction)
Migrators attracted (fraction)
0.3
0.4
0.5
0.6
0.7
0.8
0.1
0.2
0.3
0.4
0.5
5%
10%
20%
30%

In summary, effective design is both iterative and creative

Software proficiency is necessary for effective iteration…



… guided by the intellectual effort needed to construct an argument