

데이터 분석 첫 단추

데이터가 인사이트가 되려면

본 과정은, 데이터 분석이 별 것  
아니라고 느껴진다면 성공입니다.

빅데이터

데이터 분석

머신러닝

데이터 사이언스

텍스트  
마이닝

추천시스템

인공지능

하둡

NoSQL

스파크

데이터 마이닝

딥러닝

자연어 처리

빅데이터

데이터 분석

머신러닝

데이터 사이언스

스트리밍

추천시스템

하둡

인공지능

NoSQL

스파크

데이터 마이닝

딥러닝

자연어 처리



데이터에서, 뭔가

를 찾아서...



# 인사이트



# 인사이드의 정체

Users who add 7 Friends in 10 Days.

Users who do NOT add 7 friends in 10 days.





인사



인사



# 인사이드의 정체

여러분이, SNS를 운영 중이라고 해보죠!

# 인사이드의 정체

데이터 분석팀



10일 이내에 친구 7명을 사권  
사용자는 우리 서비스를 오래  
이용해주더군요!

# 인사이드의 정체

데이터 분석팀

**사실 이런 문장 자체는  
은근히 많이 발견됩니다.**

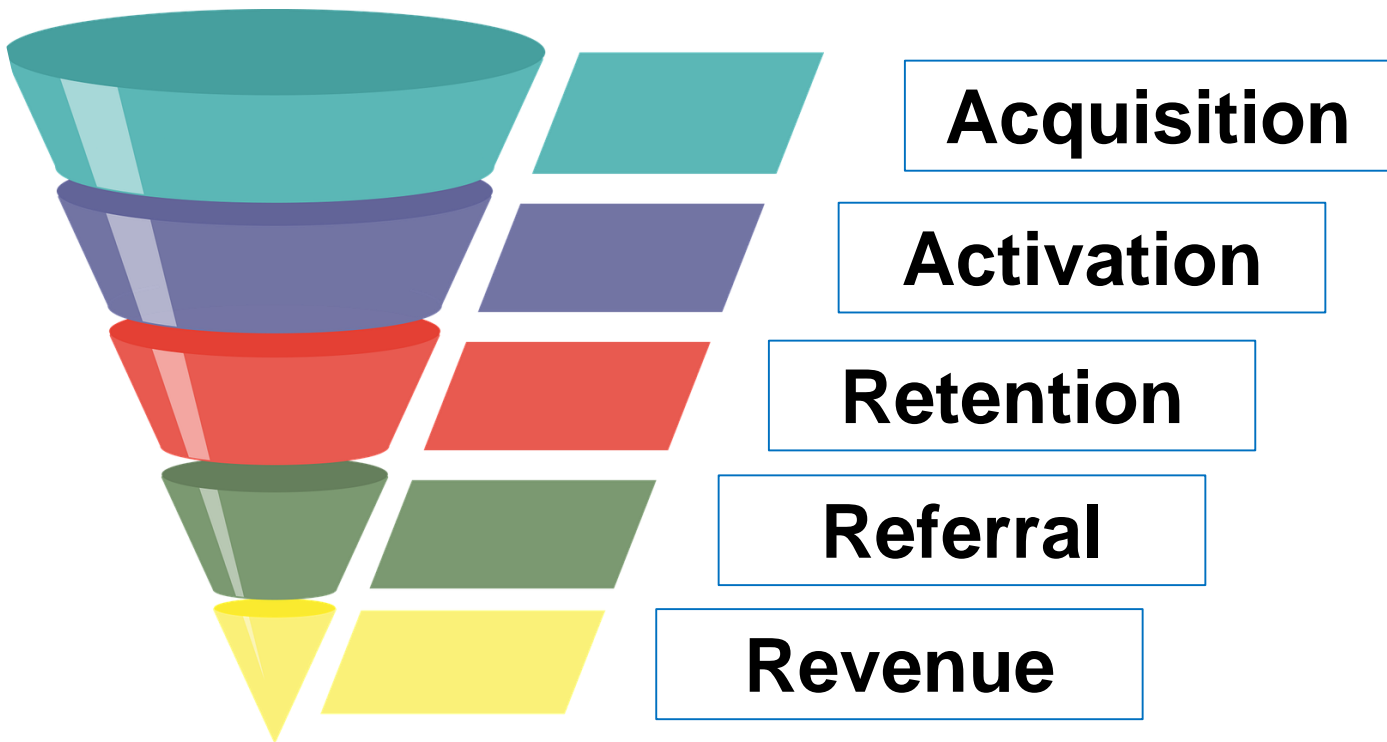
10일 이내에 친구 7명을 사귀  
는 서비스를 오래  
이용해주더군요!

인사이드의 정체

왜 페이스북에게는  
저 문장이 인사이드인가

# 인사이드의 정체

## 1. 유저 입장의 Business Model



# 인사이드의 정체

## 1. 유저 입장의 Business Model

Facebook을  
알게 됨

Facebook에  
가입함

Facebook을  
사용함



# 인사이드의 정체

## 1. 유저 입장의 Business Model

Facebook을  
알게 됨

최초의 10만 명

Facebook에  
가입함

1만 명

Facebook을  
사용함

2,000 명

# 인사이드의 정체

## 1. 유저 입장의 Business Model



# 인사이드의 정체

## 1. 유저 입장의 Business Model



# 인사이드의 정체

## 1. 유저 입장의 Business Model



하지만, 매 순간이 전부  
[관심사]

# 인사이드의 정체

## 1. 유저 입장의 Business Model



[관심사], [욕망] 중 하나  
[고객의 생존]

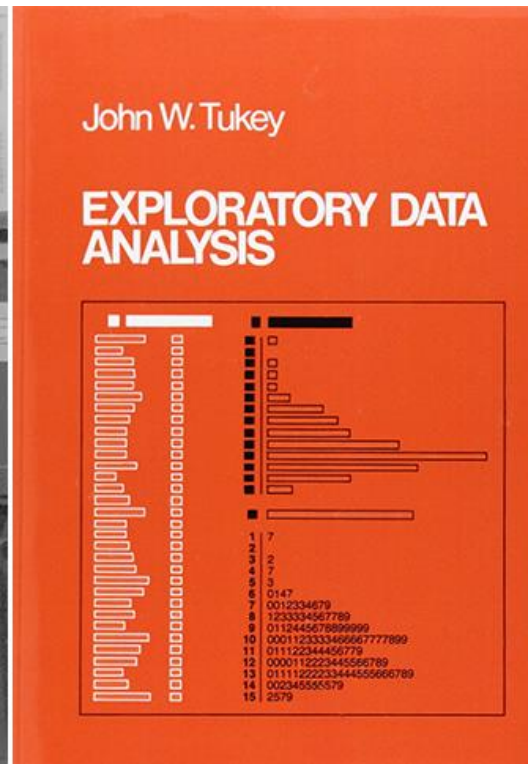
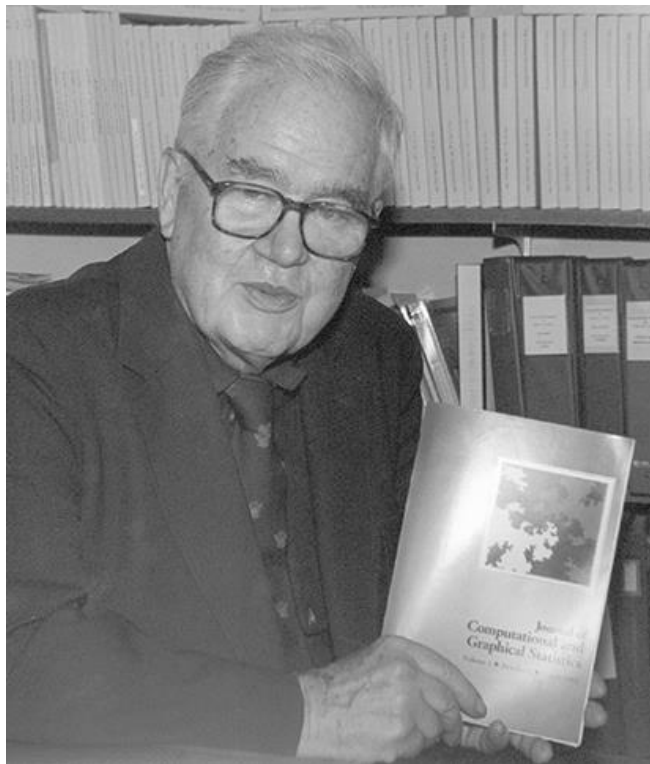
# 인사이드의 정체

## Summary

1. [욕망], [목표], [관심사]를 확실히 하라.

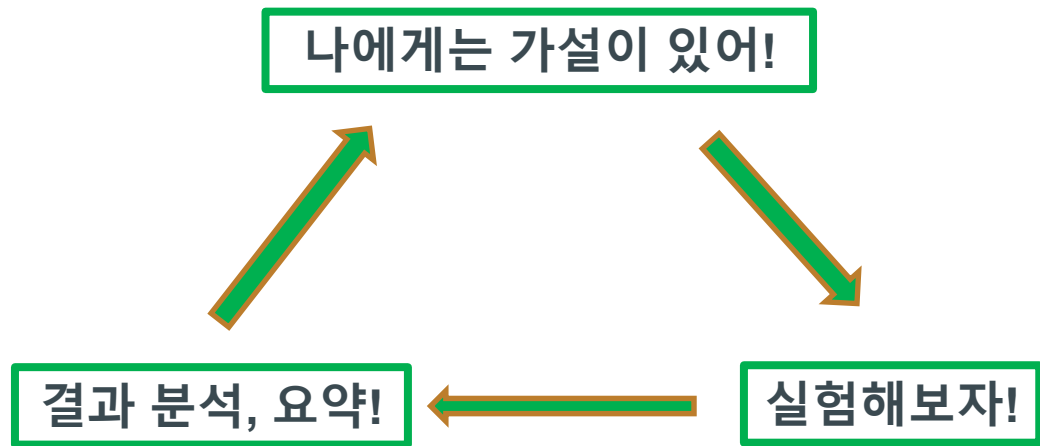
# 인사이드의 정체

## 2. 탐색하라!



# 인사이드의 정체

잠깐! [통계, 데이터 분석]이 어렵게 느껴지는 이유





# 인사이드의 정체

잠깐! [통계, 데이터 분석]이 어렵게 느껴지는 이유



나에게는 가설이 있어!

가설 검정이  
통계의 중심이 될 수 밖에  
없었음

결과 분석, 요약!

실험해보자!

# 인사이드의 정체

잠깐! [통계, 데이터 분석]이 어렵게 느껴지는 이유

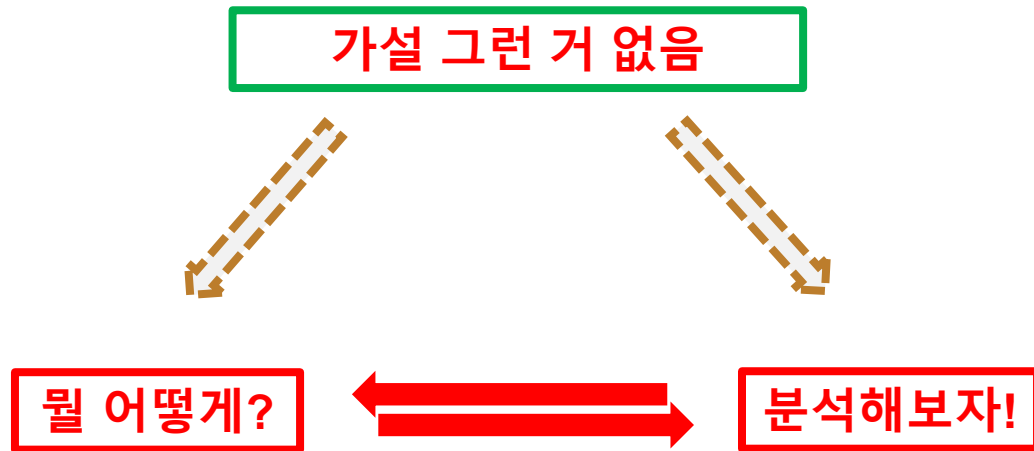
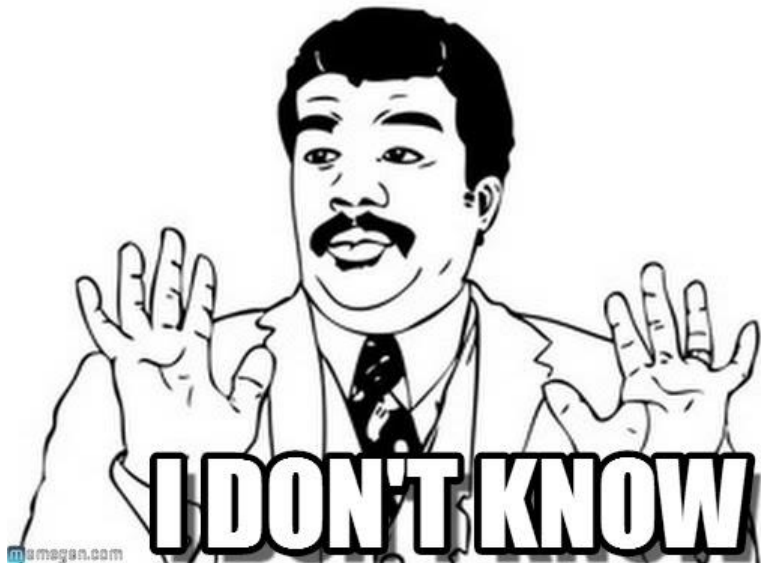


가설 넘쳐남

실험을 통해 가설을 확인하는 것이 문제  
통계는 일상

SCIENTISTS

하지만 보통은 뭘 분석 해야 할지 **모름**.  
어디서부터 시작해야 할지 **모름**.  
사실 데이터 어디 있는지도 **모름**.





## 2. 탐색하라!

[욕망], [목표], [관심사]를

**설명할 가설을 만들라**

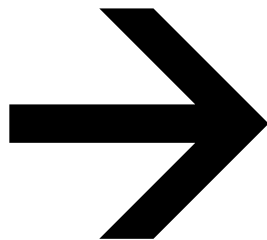
## 인사이드의 정체

### 3. [가설]과 [인사이드]의 구조

# 인사이드의 정체

## 3. [가설]과 [인사이드]의 구조

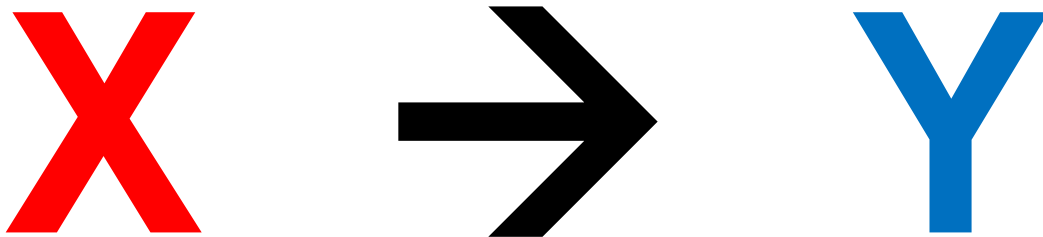
X



Y

## 인사이드의 정체

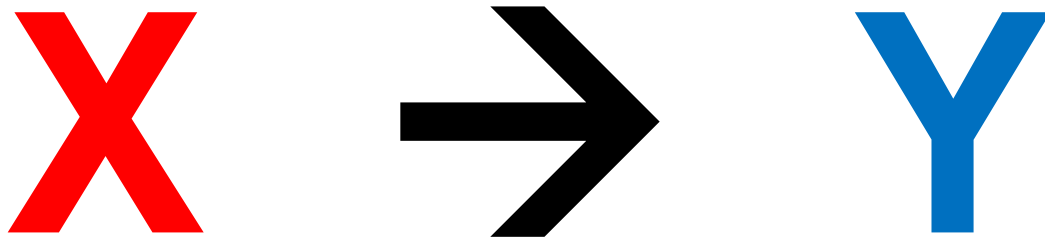
### 3. [가설]과 [인사이드]의 구조



10일 이내에 친구 7명을 사귀어 사용자는  
우리 서비스를 오래 이용해주더군요!

# 인사이트의 정체

## 3. [가설]과 [인사이트]의 구조



**X : 10일 이내 친구 7명 이상 여부**

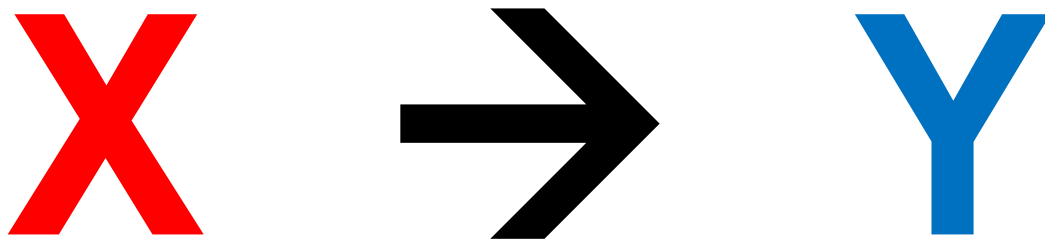
**→: 차이가 날 것이다.**

**Y : 생존율**



# 인사이드의 정체

## 3. [가설]과 [인사이드]의 구조



**X : 어제 뿌린 광고지의 개수**

**→: 비례할 것이다.**

**Y : 아이스크림 판매량**

### 3. [가설]과 [인사이드]의 구조

가설이 사실이든, 그렇지 않든

결과에 따라 행동이  
명확해져야 가설이다!

# 인사이드의 정체

## Summary

1. [욕망], [목표], [관심사]를 확실히 하라.
2. 결과에 따라 행동이 뚜렷해질 가설을 세워라
3. [가설], [인사이드]의 구조 : X → Y

## 인사이드의 정체

### 4. Again, 탐색하라!

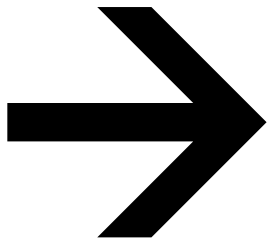


**X** → **Y**

**X를 찾아라!**  
**→ 를 찾아라!**

## 4. Again, 탐색하라!

X

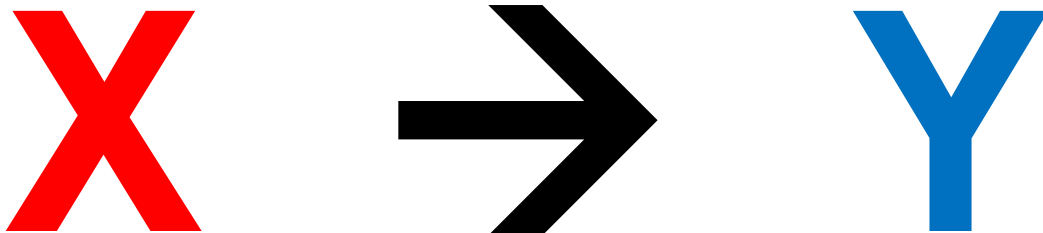


Y

고객의 생존율에...

가입요일? 광고채널? 첫 가입 후 10분 이상 사용?

## 4. Again, 탐색하라!



일단은, Data가 있고 없고는 생각하지 마세요.

# 인사이드의 정체

## Summary

1. [욕망], [목표], [관심사]를 확실히 하라.
2. 결과에 따라 행동이 뚜렷해질 가설을 세워라
3. [가설], [인사이드]의 구조 :  $X \rightarrow Y$
4. X를 찾아라 : [관심사]를 설명할 것을 찾아라

# FaceBook 에게는 인사이트일 수 밖에 없었다.



10일 이내에 친구 7명을 사권  
사용자는 우리 서비스를 오래  
이용해주더군요!



# FaceBook 에게는 인사이트일 수 밖에 없었다.



10일 이내에 친구 7명을 사권  
사용자는 우리 서비스를 오래  
이용해주더군요!

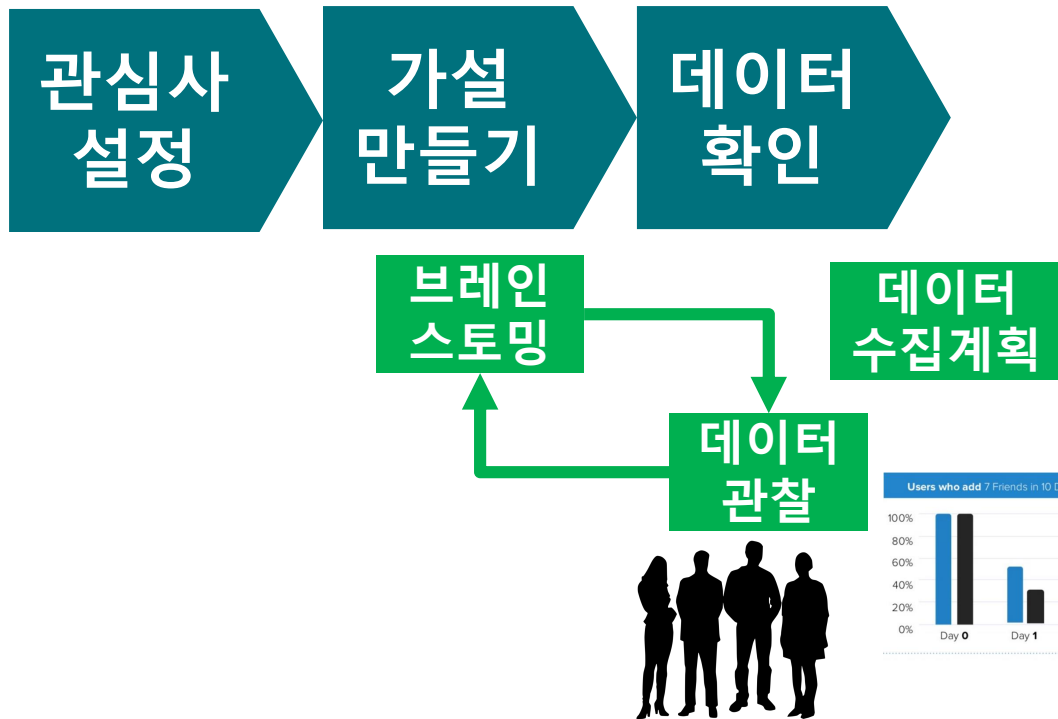
# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 아직은, 과거 사실을 기술 했을 뿐.



# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획



# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획



# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험

과거 사실을 확인하고  
가설을 만들어내는 단계

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획



과거 사실이 재현되는지  
확인하고, 가설을  
인사이트로 만드는 단계

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



우리 서비스를 오래  
사용했으면 좋겠다!

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



브레인 스토밍 : 친구가 많아야  
하나?



# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



10일 이내에 친구 7명을 사권  
사용자는 우리 서비스를 오래  
이용해주더군요! **[과거 사실]**

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



아직은 인사이트가 아님.

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



어, 새로 수집된 데이터에서도  
그러는데?

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



A : 기존 방식 그대로  
B : 친구를 억지로 추천해보자  
C : 친구를 은근히 추천해보자  
D : 친구 찾기 좋은 환경 만들자

...

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험



기존에 없던 요인이 추가 되어도,  
**어쨌든 친구를 많이 사귀게 하면  
오래오래 사용해줄까?!**

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 분석 계획

관심사  
설정

가설  
만들기

데이터  
확인

재현성  
확인

요인추가  
가설변경

실험

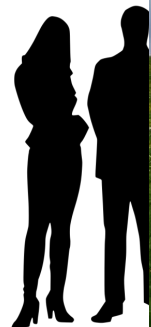


친구를 은근히 추천해주고,  
친구 찾기 쉬운 환경 만들었더니  
사용률이 올라갔어!

탐색

N

관심  
설



# FaceBook 에게는 인사이트일 수 밖에 없었다.



10일 이내에 친구 7명을 사권  
사용자는 우리 서비스를 오래  
이용해주더군요!



# 탐색적 데이터 분석 ; Exploratory Data Analysis

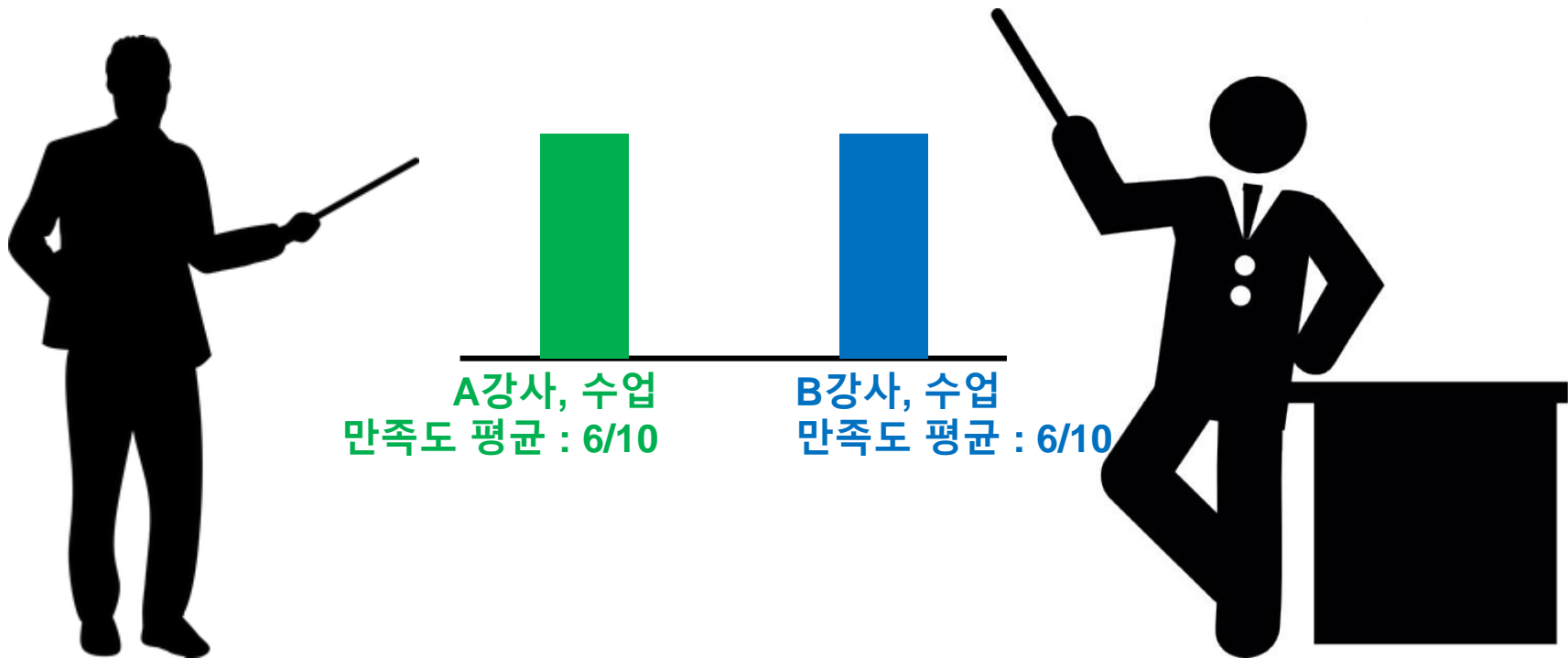
## 분석 계획 Summary

1. 뻔해 보인다고 넘기지 마라.
2. 과거에 일어난 일인지 확인하라.
3. [재현되는지] 확인하라.
4. [내가 Control할 수 있는지] 확인하라.
5. 실험하라.

# 탐색적 데이터 분석 ; Exploratory Data Analysis

방법론, 조금만 보자!

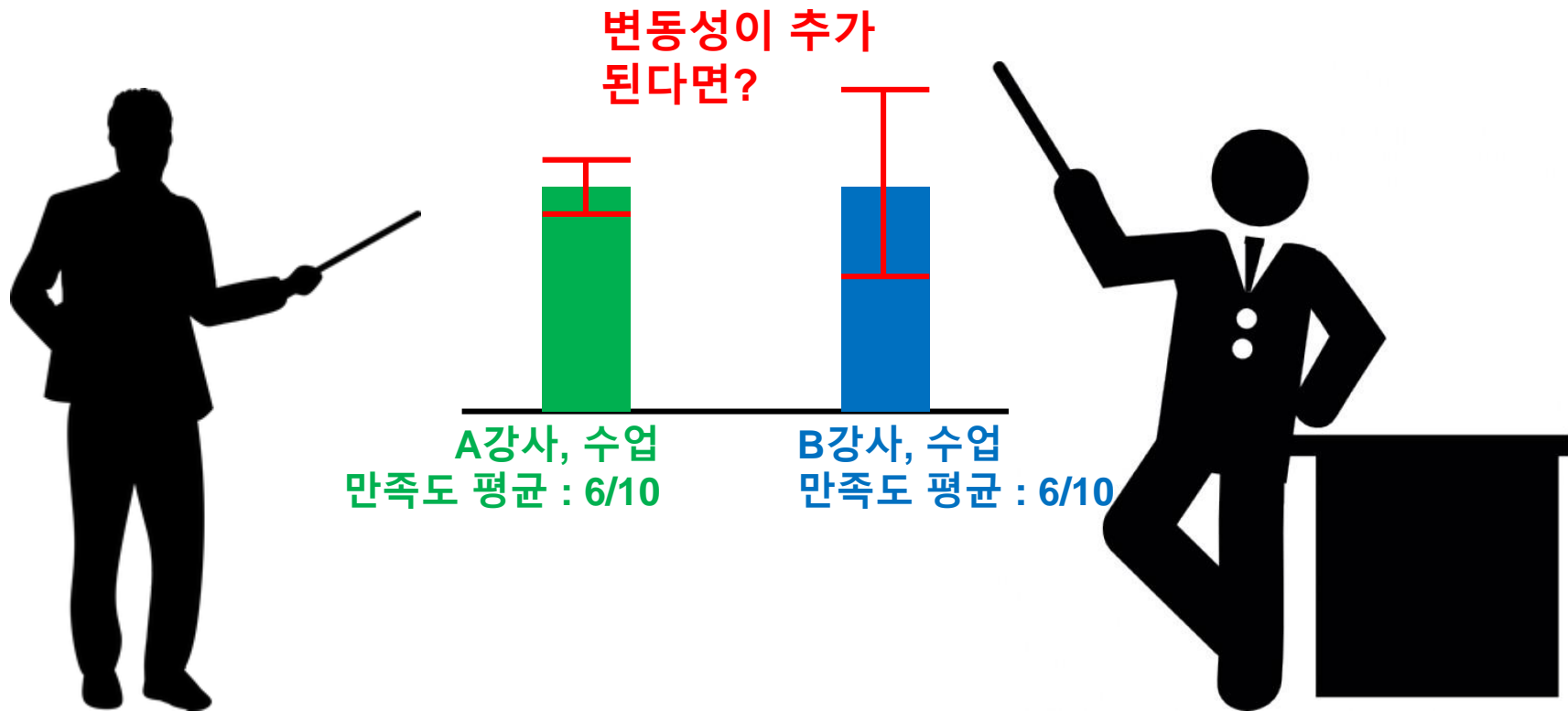
# 탐색적 데이터 분석 ; Exploratory Data Analysis



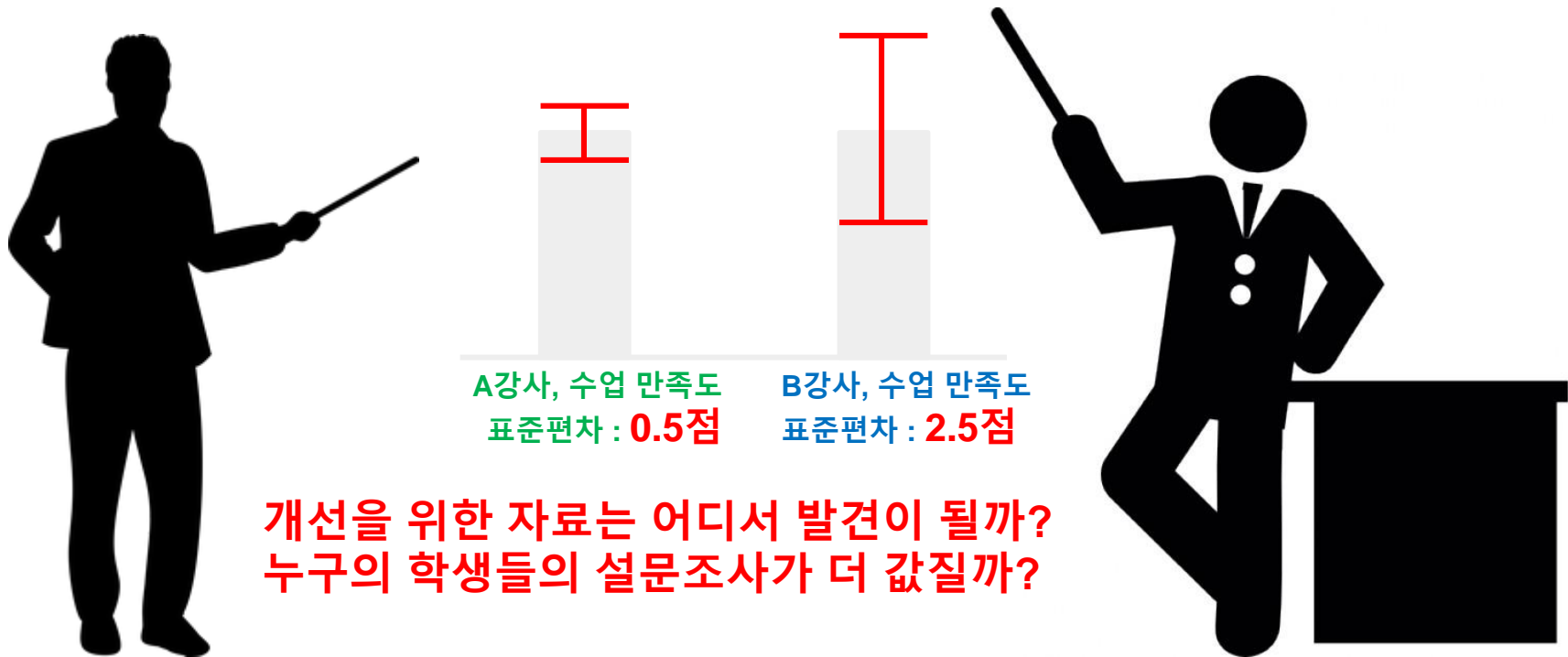
# 탐색적 데이터 분석 ; Exploratory Data Analysis



# 탐색적 데이터 분석 ; Exploratory Data Analysis



# 탐색적 데이터 분석 ; Exploratory Data Analysis



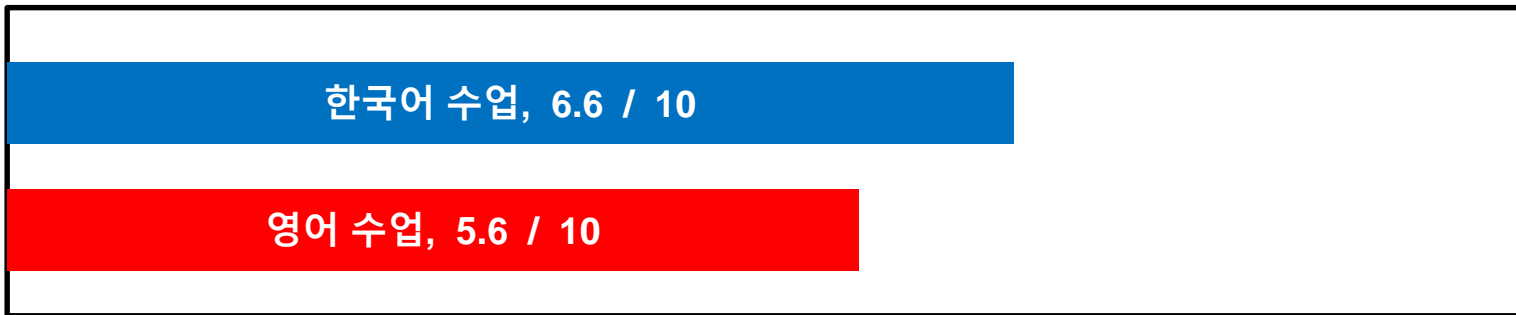
# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 방법론 Summary

1. 하나의 지표에 의존하지 말라 & 어려운 말 쓰지 말라.

# 탐색적 데이터 분석 ; Exploratory Data Analysis

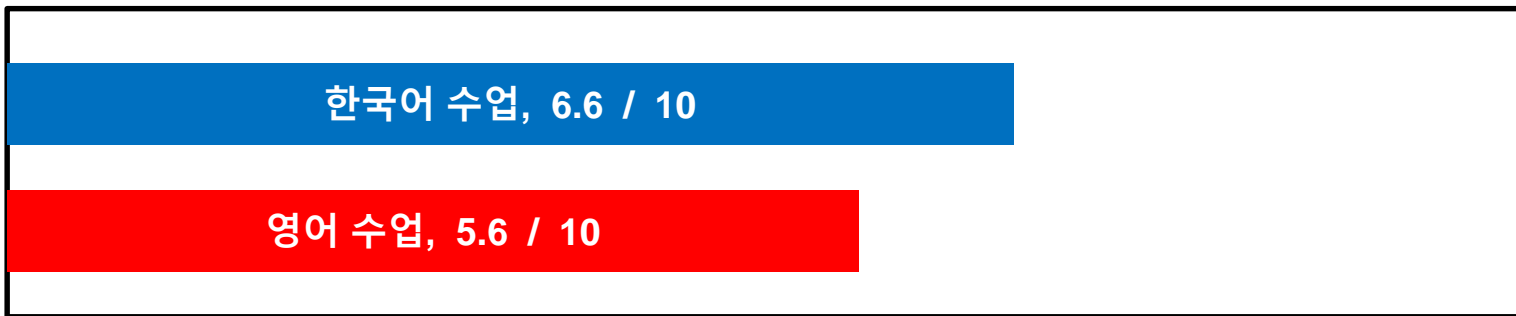
## 비교는 강력하다





# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 하지만, 쪼개야 한다.

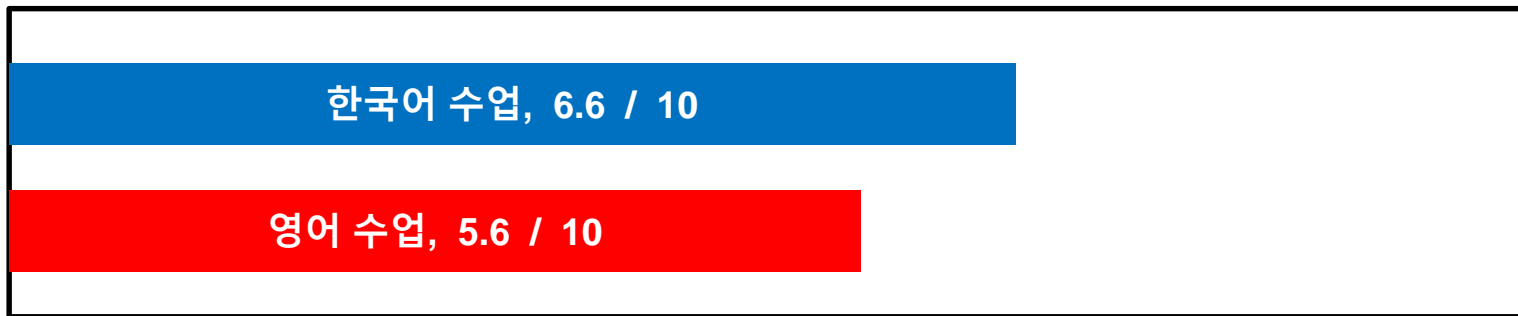


	교양	만족도 평균
한국어	90	7
영어	20	8

	전공	만족도 평균
한국어	10	3
영어	80	5

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 하지만, 쪼개야 한다.



	교양	만족도 평균
한국어	90	7
영어	20	8

	전공	만족도 평균
한국어	10	3
영어	80	5

	전체	만족도 평균
한국어	100	6.6
영어	100	5.6

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 방법론 Summary

1. 하나의 지표에 의존하지 말라 & 어려운 말 쓰지 말라.
2. 쪼개야 한다! 심슨패러독스는 수시로 일어난다!

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 비교는 강력하다!

어떤 특징이 있는 수업, 5.5 / 10

어떤 특징이 없는 수업, 4.5 / 10

# 탐색적 데이터 분석 ; Exploratory Data Analysis

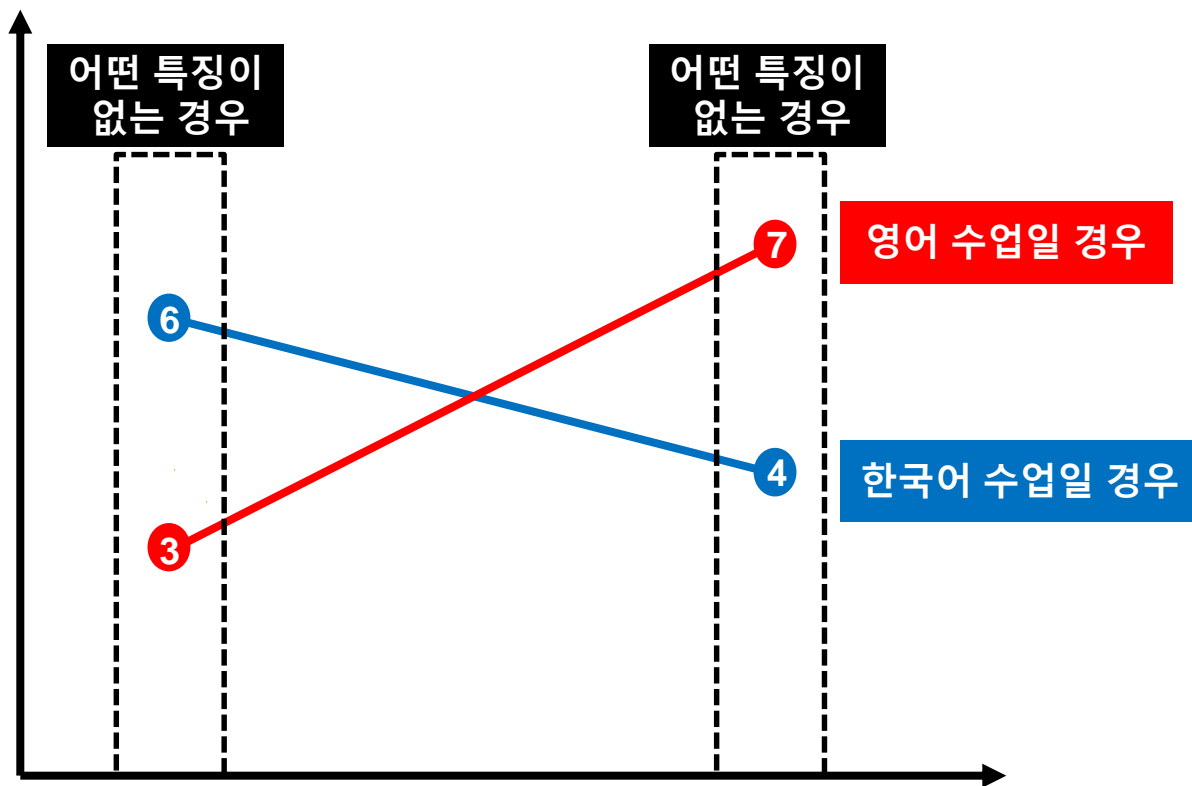
**Not Bad Case // Early Stage Case :**

어떤 특징이 없는 수업, 4.5 / 10

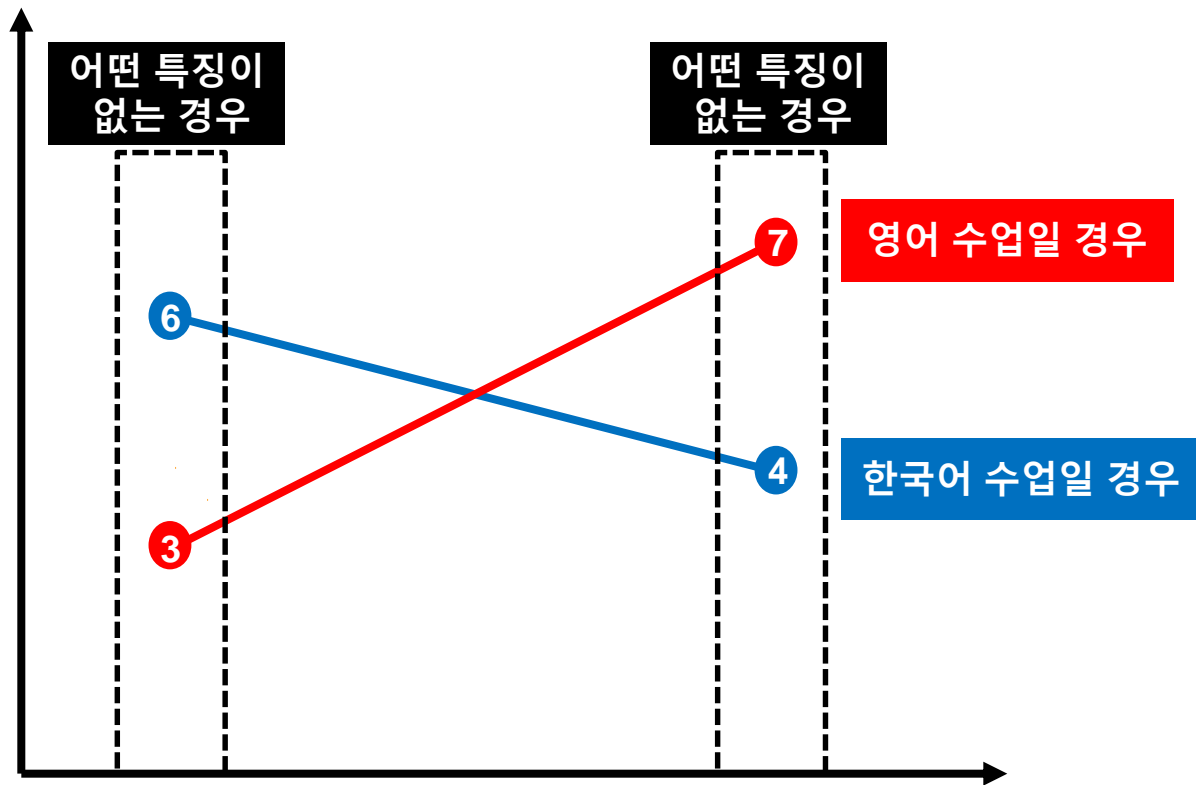
**모든 수업에 [특징]을  
실현해!**

# 탐색적 데이터 분석 ; Exploratory Data Analysis

나누자!



# 탐색적 데이터 분석 ; Exploratory Data Analysis



언어	특성	만족도 평균
한국어	있음	4
한국어	없음	6
영어	있음	7
영어	없음	3

# 탐색적 데이터 분석 ; Exploratory Data Analysis

## 방법론 Summary

1. 하나의 지표에 의존하지 말라 & 어려운 말 쓰지 말라.
2. 쪼개야 한다! 심슨패러독스는 수시로 일어난다!
3. 쪼개야 한다! 상호작용도 수시로 일어난다!



본 과정은, 데이터 분석이 별 것  
아니라고 느껴진다면 성공입니다.

# 최종 Summary, Q & A

1. 인사이트가 뭐지?
2. 데이터 → 가설 → 재현 확인 → 가설 정교화 → 실험  
→ [ 인사이트 ]
3. 지표는 관점에 따라서 다양하게.
4. 나누어 보는 것이 분석의 기본