

데이터 시각화

04. 시각화 과업과 디자인 방법

최대영 교수



고려사이버대학교
THE CYBER UNIVERSITY OF KOREA



1

● 학습리뷰

1 시각화 상호작용 개요

☞ 표현(representation) + 상호작용(interaction)

- 표현: 컴퓨터 그래픽(computer graphics) 분야
- 상호작용: 인간-컴퓨터 상호작용(HCI) 분야

☞ 두 구성요소가 기원은 다르지만 상호 밀접하게 연관되어 있음

- 사용자와 시스템의 상호작용에 의해 시각화 표현이 변경됨

2

1 시각화 상호작용 개요

☞ 시각화 상호작용 활용의 장점

- 새로운 통찰 제공
- 다른 수준의 세밀한 정보 탐색
- 데이터 중 일부를 강조
- 확장성(scalability)

1 시각화 상호작용 개요

☞ 시각화 상호작용 활용의 단점

- 상호작용이 사람이 정보를 얻는 능력을 제한
- 매체에 따라 상호작용이 제한됨
- 높은 인지부하를 가질 수 있음

2 시각화 상호작용 방법

- ✎ Select(선택하여 표시하기)
- ✎ Explore(다른 것 보여주기)
- ✎ Abstract/Elaborate(더 많이 또는 적게 보여주기)
- ✎ Reconfigure(다른 순서로 보여주기)
- ✎ Filter(일부만 보여주기)
- ✎ Encode(다른 방법으로 보여주기)
- ✎ Connect(관련 있는 것 보여주기)

5

- 시각화 과업과 과업 끝어내기에 대해 설명할 수 있다.
- 디자인 방법의 종류와 특성에 대해 설명할 수 있다.
- 파이썬의 데이터 관련 라이브러리에 대해 설명하고 이를 활용할 수 있다.

6

● 학습내용

- 1 시각화 과업과 과업 끌어내기
- 2 디자인 방법
- 3 실습(파이썬으로 데이터 다루기)



7



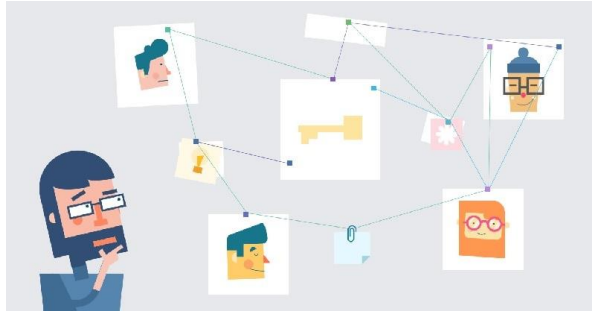
시각화 과업과 과업 끌어내기

8

1. 시각화 과업의 개념

시각화 시스템 개발에 있어 가장 중요한 일

- ☞ 사용자가 시각화를 통해서 이루려는 **목표(goal)**와 **핵심 과업(task)**이 **무엇인가를 파악**하는 것
 - 사용자의 과업을 해결하기 위한 **가장 적절한 데이터 시각화 표현** 찾기



[출처] wondershare

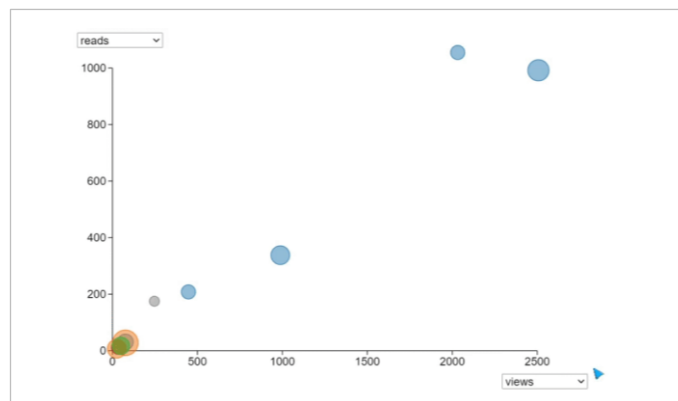
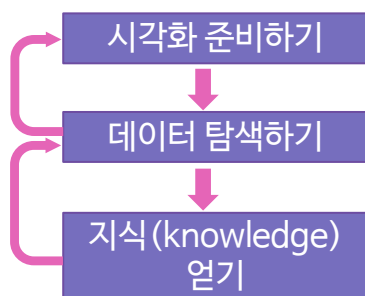
9

1. 시각화 과업의 개념

사람들이 시각화를 어떻게 사용하는가?

- ☞ 사람들이 **시각화를 어떻게 사용하는지**를 **이해**하면 사람들에게 필요한 **최적의 시각화**를 만들어내는데 도움이 됨

시각화 사용 방법



[출처] CodeX

10

1. 시각화 과업의 개념

≡ 데이터 시각화 개발자의 역할

- ☞ 사용자가 얻고자 하는 지식에 대해 이해하기
 - 어떤 종류의 지식을 만들어내려 하는가?
- ☞ 사용자가 지식을 얻을 수 있도록 시각화와 상호작용 설계하기
 - 지식을 얻는데 적절한 시각화와 상호작용은 무엇인가?



사용자가 데이터에서 지식을 획득할 수 있도록 하기 위해서는
지식을 얻기 위한 세분화된 과업(task) 필요

* 과업은 프로그래밍의 함수와 비슷한 개념

11

1. 시각화 과업의 개념

≡ 시각화 과업이란?

- ☞ 데이터를 시각적 표현으로 인코딩하여 사용자의 목적을 달성하기 위한 활동
 - 시각화는 단순히 데이터를 시각적으로 표현에 그치는 것이 아니라
사용자가 시각적 표현을 이용하여 하고자 하는 과업을 포함
- ☞ 시각화와 과업의 관계
 - “데이터 + 과업 = 시각화”
 - ➡ 어떤 시각화가 주어진 데이터에서 주어진 과업을 수행하기에 최적인가?

12

1. 시각화 과업의 개념

≡ 시각화 과업의 특징

- ☞ High level: 사용자가 왜 시각화를 사용하는가?
 - 시각화의 목적
- ☞ Middle level: 사용자가 무엇을 하는가?
 - 정보 탐색, 검색, 정리
- ☞ Low level: 사용자가 무엇을 찾는가?
 - 세부적인 데이터의 속성

13

1. 시각화 과업의 개념

≡ 데이터 시각화 개발자에게 있어서 과업의 중요성

- ☞ 개발자에게 적절한 시각화를 선택하도록 도움을 주는 것
- ☞ 과업을 고려하지 않고 데이터를 어떻게 시각화 할지를 먼저 정하면 비효율적인 결과를 만들 수 있음

14

1. 시각화 과업의 개념

≡ 과업의 예시

과학자가 두 유기체의 유전자 서열을 분석

두 유전자 서열 사이의 유사성을 찾기 위해 데이터를 탐색

서열 사이의 큰 규모의 패턴을 찾기 위한 전반적인 레벨과
작은 규모의 변화를 찾기 위한 개별적인 뉴클레오티드 사이를 옮겨다님

두 유기체의 차이를 조사하기 위해 데이터 분석의 결과를 사용

15

2. 과업 끌어내기

≡ 과업 끌어내기(Task Elicitation)란?

☞ 데이터 시각화 개발자가 사용자로부터 과업에 대한 정보를 얻어내는 것

- 시각화해야 하는 과업을 어떻게 특징지을 것인가?
- 과업의 목적, 수단, 데이터의 특징 등은 무엇인가?

☞ 과업 끌어내기의 방법

- 도메인 전문가와의 인터뷰
- 동일한 분야의 과거자료 조사

16

2. 과업 끌어내기

Why: 목적

☞ 왜 과업을 수행하는가? 사용자의 목표는 무엇인가?

- 탐색(explore): 새로운 정보를 찾기 위해 현재 정보를 탐색
- 검증(confirm): 기존 지식(prior knowledge)을 검증
- 표현(present): 효율적인 의사소통을 위해 정보를 간명하게 표현
- 시각화의 캐주얼한 사용(enjoy)도 목적의 하나

17

2. 과업 끌어내기

How: 수단

☞ 과업이 어떻게 수행되는가? 사용자와 데이터가 어떻게 상호작용하는가?

- 항해(navigation): 데이터 전반을 이리저리 옮겨 다니기
- 정리(organizing): 데이터가 표현되는 방식 변경하기
- 관계(relation): 데이터 안에서 관계성 찾기

18

2. 과업 끌어내기

What: 데이터의 특성

☞ 데이터에 대해서 무엇을 알고(learn) 싶어하는가?

- Low-level 특성: 개개의 데이터 포인트에 관심(특정 값들의 추정과 비교)
- High-level 특성: 여러 데이터 포인트들의 패턴(경향성과 평균)

☞ 시각화가 항상 사용자의 목적을 달성해주는 최상의 방법은 아님

- 사용자가 데이터의 특성을 정량화하고 싶은 경우에는 숫자 하나로 표현하는 것이 더 나을 수 있음

19

2. 과업 끌어내기

Where: 목표 데이터(Target Data)와 범위(Scope)

☞ 목표 데이터: 과업이 데이터의 어디(어느 부분)에서 수행되는가?

절대 참조 프레임 (absolute reference frame)	상대 참조 프레임 (relative reference frame)
<ul style="list-style-type: none">• 데이터를 시간, 공간과 같은 일반적인 기준과 비교	<ul style="list-style-type: none">• 다른 데이터의 비교• 인과 관계, 순서 관계 등

☞ 범위: 목표 데이터 중 얼마나 많은 부분이 과업에서 고려되는가?

- 단일 요소(세부사항), 다중 요소(문맥), 전체 요소(개괄)

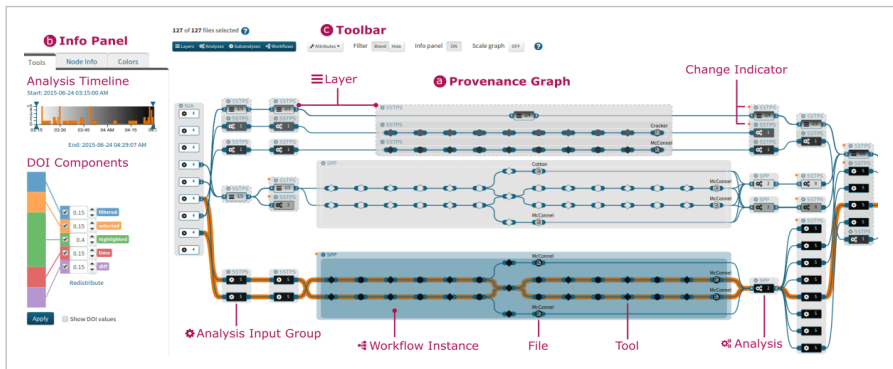
20

2. 과업 끌어내기

≡ When: 작업 흐름(Workflow)

☞ 과업은 질문과 답을 반복하는 과정으로 이루어짐 (snowballing)

- 유래 (provenance): 수행된 과업들의 히스토리 (history)
- 작업 흐름 (workflow): 과업들의 순서



21

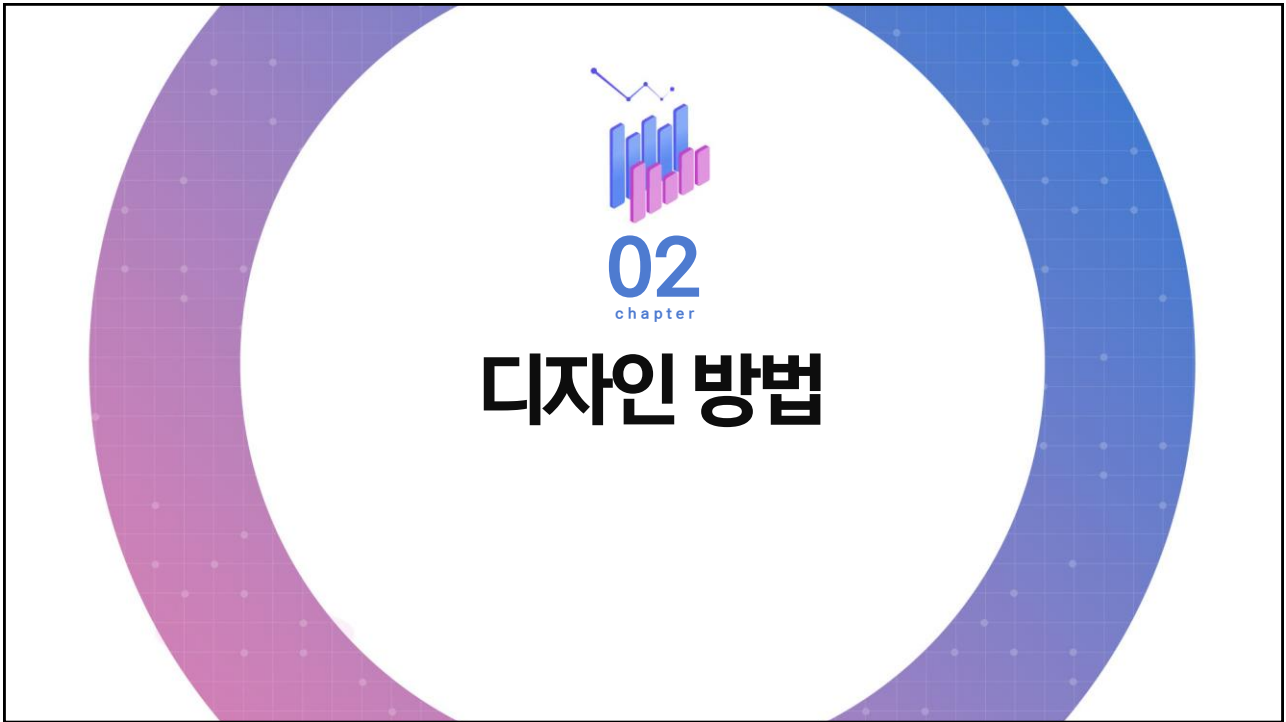
2. 과업 끌어내기

≡ Who: 역할

☞ 누가 과업을 수행하는가? 사용자는 누구인가?


- 과업의 목적과 긴밀한 관련이 있음
- 분석가, 일반적인 청중, 감독자


22




23

1. Five Design Sheet

 시각화 디자인 프로젝트

 User-Centered Design Process(사용자 중심 디자인)

- 디자이너(데이터 시각화 개발자)는 사용자가 원하는 시각화를 디자인하기 위해 사용자와 긴밀히 협력
- 사용자가 디자인의 전과정에서 주기적으로 피드백을 제공

 사용자와 함께 협력하여 3가지 요소를 파악

1

 사용자가 가진 데이터(data abstraction)

2

 사용자가 가진 질문(task analysis)

3

 시각화하기 위한 전략: Five Design Sheet, Design Study Methodology

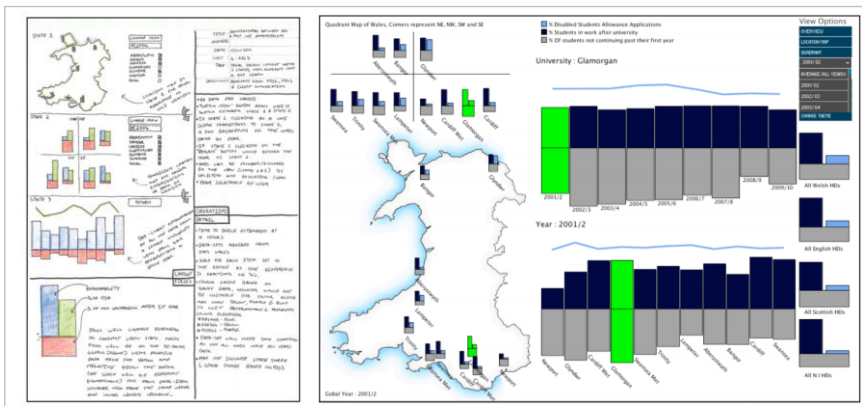
24

1. Five Design Sheet

프로토타입 만들기(Prototyping)

프로토타입을 만들고 **계속적인 반복을 통해 구체화해 나감**

- 계속해서 비판적으로 검토하고 사용자의 동의를 얻어 수정



[출처] Old Dominion University CS725

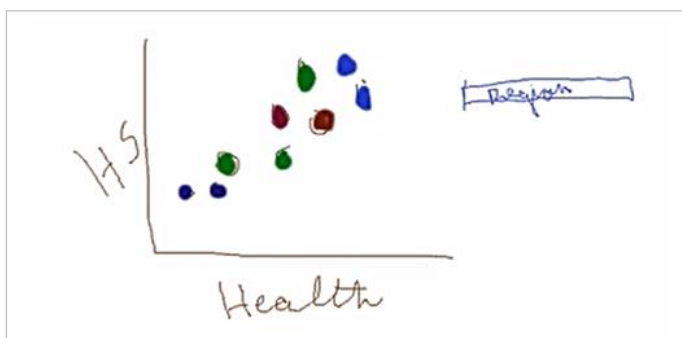
25

1. Five Design Sheet

Prototyping의 수준

저해상도 (low-fidelity) 프로토타입

- 빠른 아이디어내기(ideation)와 사용자 피드백을 받기 위한 낮은 수준의 기술(스케치 등)을 이용한 프로토타입



[출처] Fundamentals of Data Visualization, University of Colorado Boulder

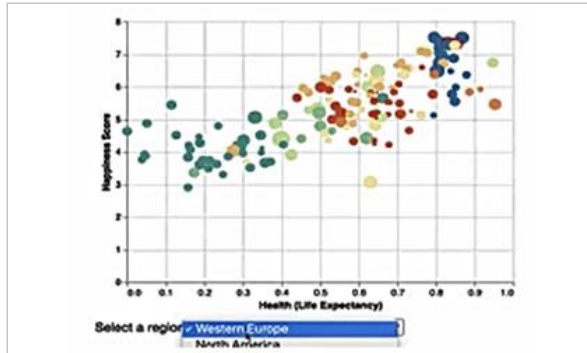
26

1. Five Design Sheet

≡ Prototyping의 수준

☞ 고해상도 (high-fidelity) 프로토타입

- 정제되고 기능이 최종 시각화 완성품과 유사한, 마지막 사용자 피드백을 받기위한 단계에서의 프로토타입



[출처] Fundamentals of Data Visualization, University of Colorado Boulder

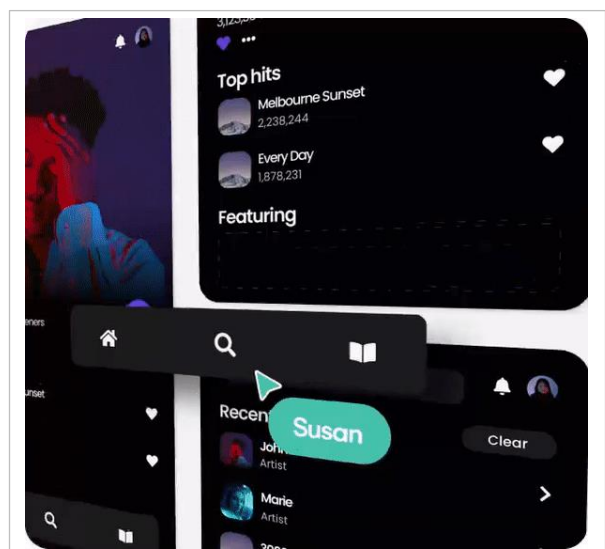
27

1. Five Design Sheet

≡ Prototyping의 수준

☞ Prototyping 도구(tool)

- 시각화의 매체 종류에 따라 prototyping 전용 도구를 이용하여 저해상도와 고해상도의 장점 조합



[출처] uizard

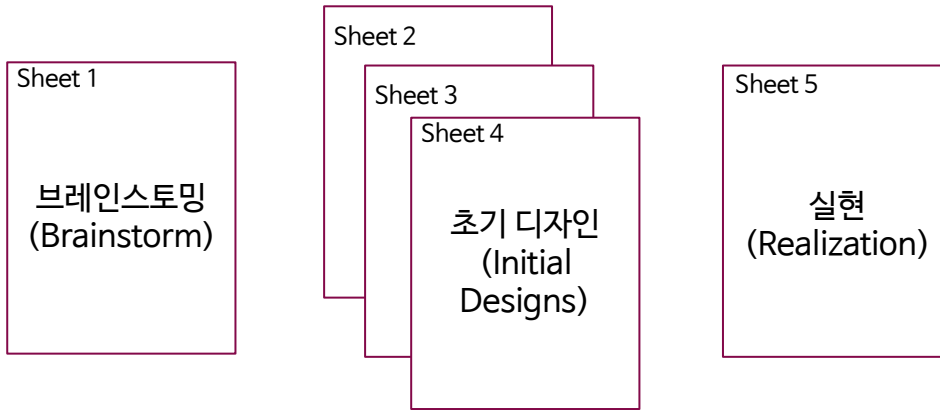
28

1. Five Design Sheet

≡ Five Design Sheet 방법

☞ 사용자(이해관계자)와 긴밀히 협력하여 모형을 만드는 방법 중 하나

- 5장의 sheet를 만들어 이해관계자와 의사소통



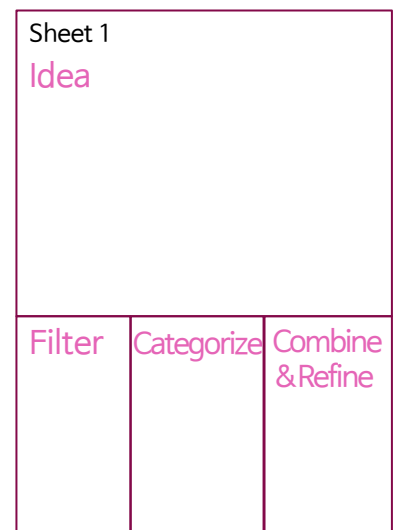
29

1. Five Design Sheet

≡ Five Design Sheet 방법

☞ Sheet 1: Brainstorm

- 아이디어(idea): 가능한한 많은 아이디어 내기
- 선별(filter): 중복, 불가능한 아이디어 제거하기
- 범주화(categorize): 관련 있는 아이디어 묶기
- 조합과 정제(combine & refine): 좋은 아이디어 조합하고 다듬기



30

1. Five Design Sheet

≡ Five Design Sheet 방법

📌 Sheet 2~4: Initial Design

- 레이아웃(layout): 최종 시각화의 모습을 스크린샷처럼 표현한 전체모습 나타내기
- 중심(focus): 시각화의 주요 아이디어를 세부적으로 나타내기(특정 부분이나 상호작용)
- 작동(operations): 사용자가 시각화를 어떻게 사용하고 상호작용하는지 나타내기
- 논의(discussion): 적용한 기술의 장단점 등을 분석하기

Sheet 2, 3, 4 Layout	Administrivia
	Operations
Focus	Discussion

31

1. Five Design Sheet

≡ Five Design Sheet 방법

📌 Sheet 5: Realization

- Layout, focus, operations 부분을 완성하기
- Sheet 2, 3, 4에서 수정하거나 조합을 해도 됨
- Discussion에는 다음의 자세한 정보 기입
 - 알고리즘
 - 디자인 세부사항(색, 형태 등)
 - 예상 작업시간
 - 하드웨어/소프트웨어 요구사항

Sheet 5 Layout	Administrivia
	Operations
Focus	Discussion (Details)

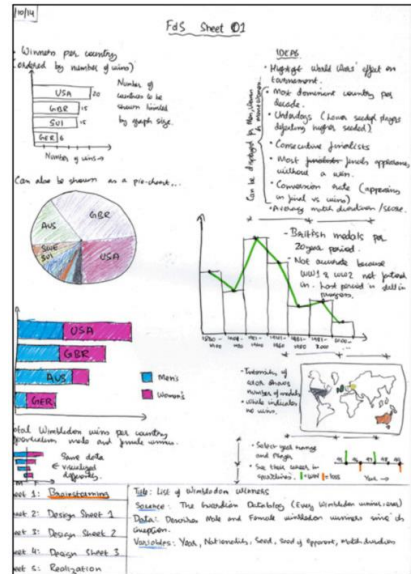
32

1. Five Design Sheet

Five Design Sheet 예시

Sheet 1: Brainstorm

- 올림픽 메달 수 관련 분석 데이터
- 막대 그래프, 파이차트
- 세계2차 대전 동안의 자료 없음



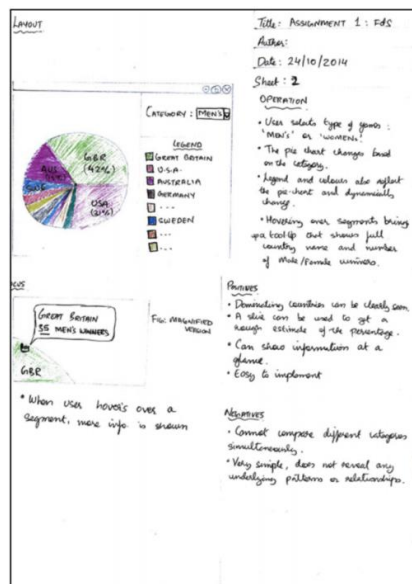
[출처] Old Dominion University CS725

33

1. Five Design Sheet

Five Design Sheet 예시

Sheet 2: Initial Design 1



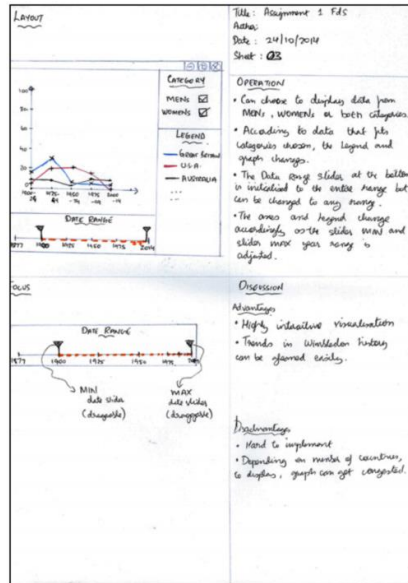
[출처] Old Dominion University CS725

34

1. Five Design Sheet

Five Design Sheet 예시

Sheet 3: Initial Design 2



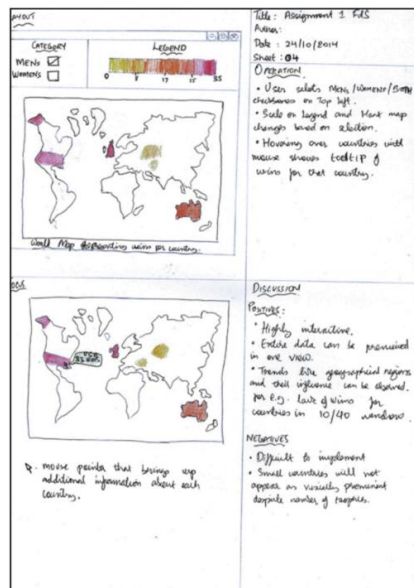
[출처] Old Dominion University CS725

35

1. Five Design Sheet

Five Design Sheet 예시

Sheet 4: Initial Design 3



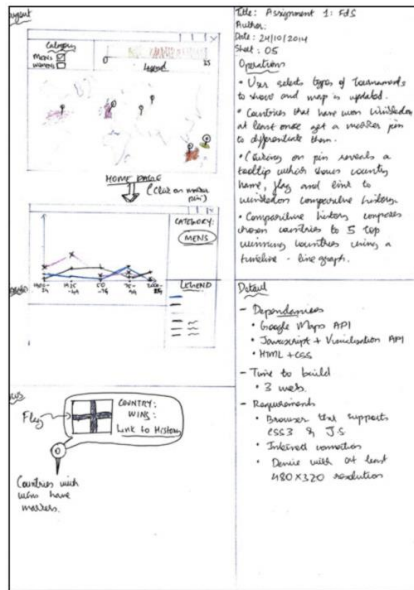
[출처] Old Dominion University CS725

36

1. Five Design Sheet

Five Design Sheet 예시

Sheet 5: Realization



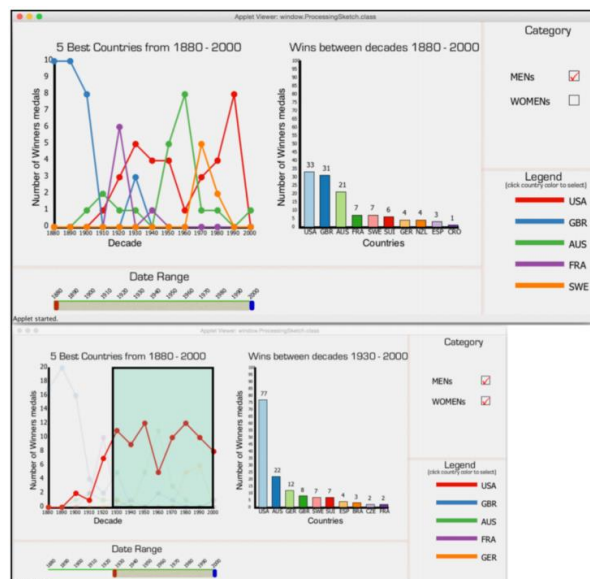
[출처] Old Dominion University CS725

37

1. Five Design Sheet

Five Design Sheet 예시

시각화 결과



[출처] Old Dominion University CS725

38

2. Design Study

≡ 스케치 기반 prototyping의 장·단점

장점	단점
<ul style="list-style-type: none">• 비용이 적게 들고 작업이 빨라 여러 번의 반복이 가능• 사용자와 의사소통하기에 매우 편함	<ul style="list-style-type: none">• 추측에 기반하여 짧은 시간 내에 만들어진 프로토타입은 실제로 시각화하기 전에는 그 성공 여부를 알 수 없음

39

2. Design Study

≡ Design Study

A **project** that **analyzes** a real-world problem in order to **design** a visualization system that contains a **validated** design where designers **reflect** about lessons learned

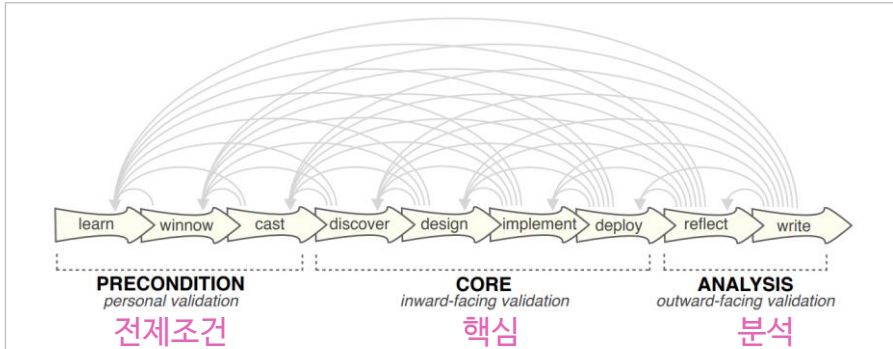
현실 데이터를 이용하여 실제 문제를 **분석**하는 **프로젝트**로
디자이너가 이해관계자와 의사소통하고 협력하여
검증된 디자인이 포함된 시각화 시스템을 **디자인**하고
프로젝트 이후에는 프로젝트에서 얻은 경험과 지식을 **활용**

40

2. Design Study

Design Study Methodology(디자인 연구 방법)

- ☑ 9단계가 순서대로 진행되고 이전 단계로 돌아가 반복 가능
- ☑ 앞 단계를 바로 건너뛰는 것은 제한
 - 구현(implement) 없이 설치(deploy)는 불가



[출처] Sedlmair et al., Design Study Methodology: Reflections from the Trenches and the Stacks

41

2. Design Study

전제조건(Precondition)

- ☑ Learn(문헌 연구하기)
 - 시각화할 도메인(분야)에 대한 문헌을 읽고 전문가와 논의
 - 도메인에 대한 지식을 쌓는 것은 사용자와의 의사소통을 위해 필수
- ☑ Winnow(적절한 공동작업자 찾기)
 - 디자이너와 긴밀히 협력할 유망한 도메인 전문가 선정
- ☑ Cast(공동작업자의 역할 식별하기)
 - 주분석가(front-line analyst), 문지기(gate keeper), 협력디자이너(co-designer) 등

42

2. Design Study

≡ 핵심(Core)

☞ Discover(문제 특징짓기와 추상화하기)

- 어떤 과업을 풀려고 하는가?
- 가지고 있는 데이터는 무엇인가?

☞ Design(데이터 추상화하기, 시각화 표현과 상호작용 만들기)

- 필요한 데이터를 추출하고 변형하는 추상화 과정을 포함
- Five Design Sheet 방법 등 스케치 방법을 이용 시제품 만들기

43

2. Design Study

≡ 핵심(Core)

☞ Implement(프로토타입 또는 시스템 만들기)

- 시각화를 위한 적절한 소프트웨어 알고리즘과 도구를 선정
- 실제 데이터를 가지고 고해상도 시각화 구현

☞ Deploy(사용자에게 시각화를 제공하고 피드백 받기)

- 사용자가 새로운 시각화의 도움을 받는지를 확인
- 사용자가 업무와 문제를 보는 방식을 변화시키는지를 확인

44

2. Design Study

≡ 분석(Analysis)

🔗 Reflect(프로젝트 되돌아보기)

- 시각화 디자인과 도메인에 대한 가설 확정 또는 개선
- 프로젝트 이전에 몰랐던 것 중에 새로 알게 된 것은 무엇인가?

🔗 Write(프로젝트로 부터 알게 된 것을 요약하기)

- 프로젝트를 진행하는 과정, 디자인 요소, 도메인 지식 등을 기술
- 프로젝트에 대한 문서화, 블로그 포스트 등을 작성

45

2. Design Study

≡ Design Study의 특징

01 Task-centered(과업 중심)

- 핵심 질문과 목표를 중심으로 프로젝트 진행

02 User-centered(사용자 중심)

- 프로젝트의 전과정에서 사용자(이해관계자)와 협력

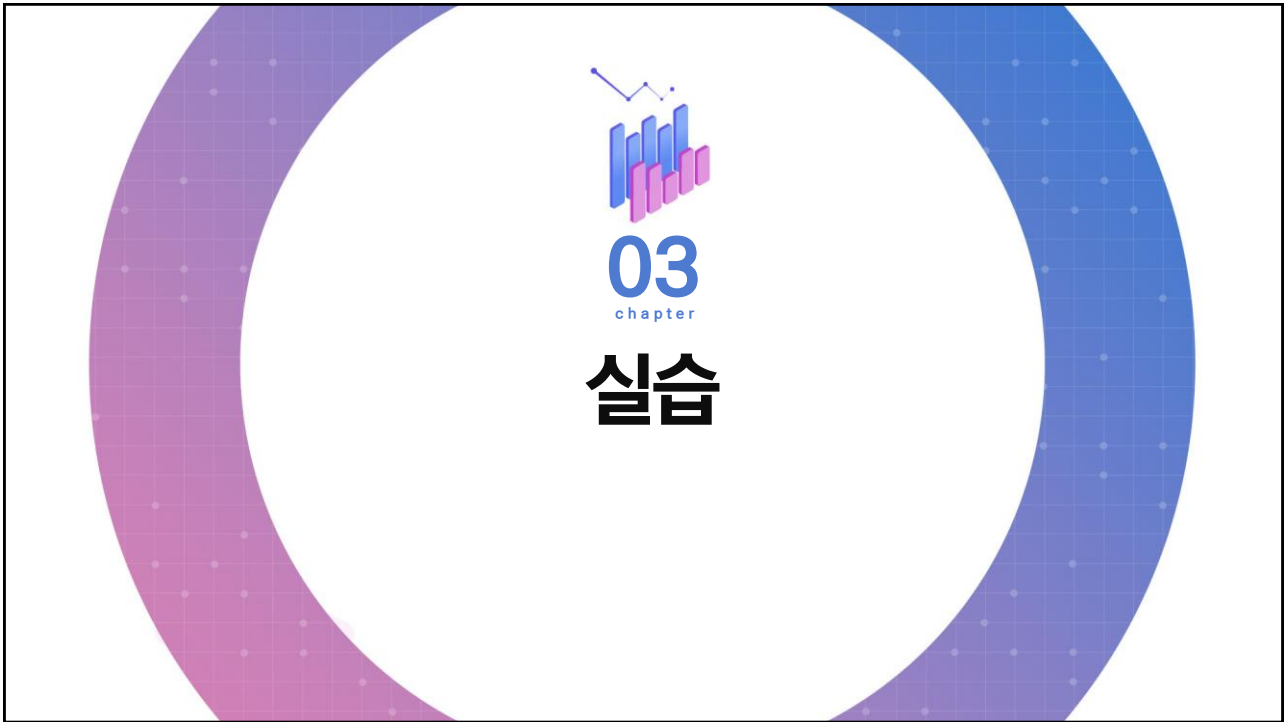
03 Iterative(반복)

- 이해관계자의 의견이 수렴할 때까지 충분히 핵심 단계를 반복

04 Evaluative(평가)

- 문제를 풀면서 이해관계자들이 얻은 통찰을 함께 평가

46



1. 파이썬으로 데이터 다루기

파이썬 실습환경

Colab (Colaboratory)

구글에서 교육과 연구 목적으로 무료로 제공

파이썬 코드 작성과 수행이 가능한 환경 제공 (웹브라우저에서 실행)

개발환경 설정 불필요 및 하드웨어 가속기(GPU, TPU) 지원

작성된 코드를 구글 드라이브에 저장 가능(간편한 공유)

CO

Colab (Colaboratory)

구글에서 교육과 연구 목적으로 무료로 제공

파이썬 코드 작성과 수행이 가능한 환경 제공 (웹브라우저에서 실행)

개발환경 설정 불필요 및 하드웨어 가속기(GPU, TPU) 지원

작성된 코드를 구글 드라이브에 저장 가능(간편한 공유)

출처] <https://colab.research.google.com/?hl=ko>

48


24

1. 파이썬으로 데이터 다루기

데이터 관련 라이브러리

📄 pandas

- 행과 열로 이루어진 데이터를 처리하는 다양한 기능을 제공
- 1차원 데이터구조(Series)와 2차원 데이터구조(DataFrame)로 구성
- 엑셀 대비 대용량 데이터 처리에 효과적
- 문자, 숫자 등 다양한 자료형을 데이터 처리 가능



```
In [1]: import pandas as pd
In [2]: students = pd.read_csv("students.csv")
students.head()
Out[2]:
```

	Name	RollNo	Date Of Admission	Emergency Contact
0	Shubham	1	20-05-2012	9988776655
1	Gagan	2	20-05-2009	8364517829
2	Oshima	3	20-05-2003	5454223344
3	Vyom	4	20-05-2009	1223344556
4	Ankur	5	20-05-1999	9988776655

[출처] <https://pandas.pydata.org/>


49

1. 파이썬으로 데이터 다루기

데이터 관련 라이브러리

📄 NumPy

- 다차원 배열, 배열 간 연산, 정렬 등 데이터를 처리하는 다양한 기능을 제공
- 일반 리스트에 비해 빠르고 메모리 공간을 적게 차지함
- 동일한 자료형만 담을 수 있음(리스트는 다양한 자료형 가능)
- 컴퓨터 비전, 언어 모델 등 대량의 숫자를 다루는데 쓰임



```
>>> data = np.array([[1, 2], [3, 4], [5, 6]])
>>> data
array([[1, 2],
       [3, 4],
       [5, 6]])
```

np.array([[1, 2], [3, 4], [5, 6]]) →

1	2
3	4
5	6

[출처] <https://numpy.org/>

50

● 학습정리

1 시각화 과업과 과업 끌어내기

📌 시각화 과업

- 데이터를 시각적 표현으로 인코딩하여 사용자의 목적을 달성하기 위한 활동
- 시각화는 단순히 데이터를 시각적으로 표현에 그치는 것이 아니라 사용자가 시각적 표현을 이용하여 하고자 하는 과업을 포함

📌 과업 끌어내기

- 데이터 시각화 개발자가 사용자로부터 과업에 대한 정보를 얻어내는 것
- 목적, 수단, 데이터의 특성, 목표 데이터와 범위, 작업 흐름, 역할

51

● 학습정리

2 디자인 방법

📌 Five Design Sheet 방법

- 5장의 sheet를 만들어 이해관계자와 긴밀히 협력하여 시각화 모형을 만드는 방법

📌 Design Study Methodology

- 전제조건, 핵심, 분석의 3개의 범주로 구분
- 9단계로 구분되며 순서대로 진행되고 이전 단계로 돌아가 반복 가능

52

● 학습정리

3 실습(파이썬으로 데이터 다루기)

📁 pandas

- 행과 열로 이루어진 데이터를 처리하는데 주로 쓰이는 라이브러리

📁 NumPy

- 다차원 배열의 연산, 정렬 등을 처리하는데 주로 쓰이는 라이브러리

53

● 참고문헌

- 📖 「Fundamentals of Data Visualization」, Danielle Albers Szair, University of Colorado Boulder(coursera course).
- 📖 「Design Study Methodology: Reflections from the Trenches and the Stacks」, Sedlmair et al., IEEE Transactions on Visualization and Computer Graphics, 2012.

※ 서체 출처 | 넥슨Lv2고딕-(넥슨코리아)www.levelup.nexon.com / 나눔바른고딕(네이버)

54

저작권 안내

이 강의록은 저작권법에 의해 보호받는 저작물로서
저작권자의 허락 없이 저작재산권 일체(복제권,
배포권, 대여권, 공연권, 공중전송권, 전시권, 2차적
저작물 작성권)를 침해 시 저작권법에 의거 처벌받을
수 있습니다.