VIIII. 효과적인 시각화 만들기

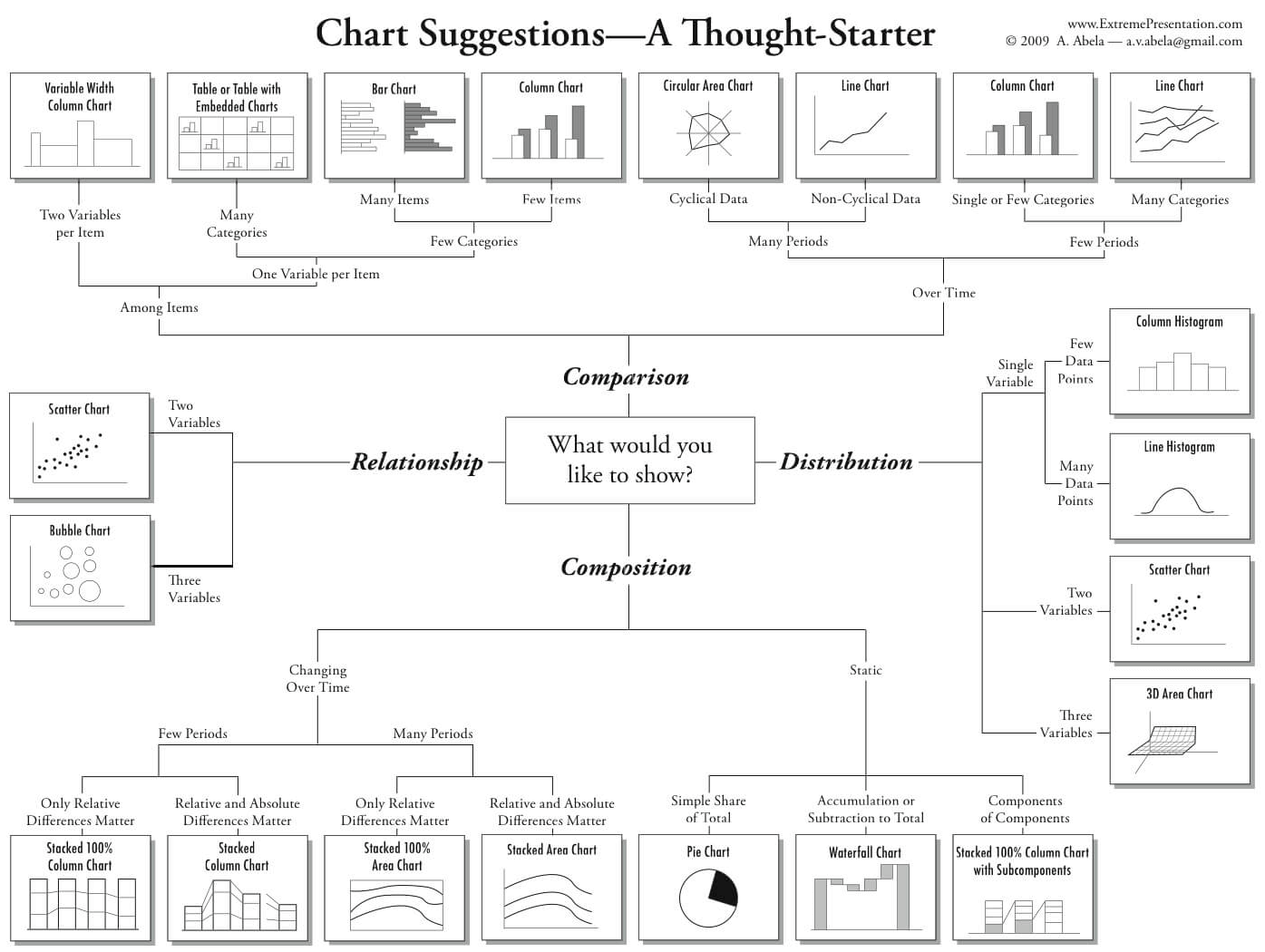
# 시각화의 목적 : 스토리텔링

R이든 python이든, 어떤 패키지나 라이브러리를 쓰든 데이터 시각화를 만드는 사용자들은 가끔, 아니 자주 시각화의 테크닉을 최대한 사용하여 수려한 그래프를 만들기 위해 노력한다. 그렇게 많은 기술적 요소를 가진 시각화를 만드는 사용자는 자신의 기술에 대해 뿌듯할지 모르지만, 많은 경우 화려한 기술에 데이터가, 데이터가 말하는 의미가 퇴색되는 경우가 종종 발생한다. 그래서 시각화 디자인을 만들기 전에는 반드시 그 데이터가, 시각화가 무엇을 의미하고, 무엇을 전달하려 하는지를 명확히 해야한다. 이를 스토리텔링이라고 한다.

특히 최근에는 데이터 분석의 결과를 대중에게 전달할 때 이 스토리텔링을 중요하게 여긴다. 이것을 영문으로 KWYRWTS라고 표시하는데 이는 ‘Know What You Really Want To Say’(네가 진정 이야기하고자 하는 것을 알아라)의 준말이다.

# 적절한 시각화 타입의 선정

시각화의 목적이 설정되었다면 다음은 어떤 시각화 타입으로 시각화를 만들지를 선택해야 한다. 이 책에서는 시각화를 하는 목적에 따라 시각화를 설명하였다. 데이터의 관계성을 보여줄지, 데이터 간의 비교를 보여줄지, 데이터의 추세를 보여줄지, 데이터의 분포를 보여줄지 등의 목적에 따라 선택할 수 도 있고, 그 데이터의 종류가 두개의 연속형 데이터인지, 하나의 연속셩 데이터와 하나의 이산형 데이터인지, 두 개의 이산형 데이터인지, 세 가지 이상의 데이터를 표현해야 할지와 같이 그 데이터의 종류에 따라 결정할 수도 있다. 다음의 그림은 시각화의 목적에 따른 차트의



이 스토리텔링이 대중에게 잘 전달되기 위해서는 다음과 같은 것들을 고려해야 한다.

# 간결한 시각화

시각화는 최대한 간결하게 그려져야 한다. 그려야 하는 시각화의 요소들이 과연 데이터를 설명하기 위해 필요한지, 데이터에 대한 이야기를 이어가는데 필요한지를 먼저 고려해야 한다. 만약 그 필요성이 명확치 않다면 과감하게 제거해야 한다. 특히 불필요한 텍스트, 과도한 색상, 어지러운 그림자, 3D 이미지로 시각화를 어지럽히지 않도록 주의해야 한다.

시각화는 대부분 그래픽적 도형으로 구성되지만 그 도형들을 빠르게 이해시키기 위해서는 적절한 곳에 텍스트 레이블을 사용한다. 앞서 5장에서 살펴본 바와 같이 범례는 데이터를 구분하기 위해 매우 중요한 요소임에는 틀림없지만 볌례를 확인하기 어렵거나 인식하는데 너무 많은 시간이 걸린다면 적절한 선이나 막대 옆에 레이블을 배치하는 것도 좋은 선택일 수 있다.

그리드 눈금선은 데이터를 더 쉽게 해석할 수 있도록 사용되어야 한다. 차트가 명확하고 간결한 느낌이 들도록 너무 많지 않게, 또 데이터의 해석이 가능하게 너무 적지 않게 설정하여야 한다.

시각화에는 적절한 여백이 필요하다. 사람의 눈은 너무 많은 디자인 요소를 해석하는데 익숙하지 못하다.

디자이너는 적을수록 좋다는 것을 직관적으로 알고 있습니다. 여백을 받아들입니다. 눈은 모든 콘텐츠 사이에 들어갈 장소가 필요합니다. 청중을 압도하지 마십시오.

이 단순한 시각적 미학에 대한 많은 논쟁이 있습니다. 미리 주의를 기울이는 청중보다 더 중요한 것은 없습니다. 내가 말했듯이, 누군가가 당신의 데이터 시각화를 보는 순간, 그들은 그것을 판단하고 있는 것입니다. 그 순간에 (무의식적으로) 그들은 스스로에게 무엇을 말하고 있습니까?

“무엇이 가장 중요합니까?” “여기서 무엇을 봐야 합니까?” “무엇을 무시해야 합니까?” “이게 나에게 너무 복잡한가요?” “이게 너무 길어서 시간을 할애할 가치가 없나요?” “이게 믿을만한 것 같습니까?” 그리고 더… 연구에 따르면 인포그래픽 디자인에서 가장 중요한 요소 중 하나는 청중의 주변 시야를 이해하는 것입니다. 즉, 사용자가 한 눈에 볼 수 있는 것이 다음에 볼 위치를 결정한다는 것입니다. 즉, 사전 주의 깊은 방식으로 시각화의 특정 지점으로 그들을 끌어들일 뿐만 아니라 다음에 볼 무언가를 제공해야 한다는 것을 인식해야 합니다!

이 아이디어를 설명하기 위해 생각할 수 있는 가장 좋은 예는 지도 인터페이스를 생각하는 것입니다. 레스토랑의 위치와 순위를 매핑하고 있습니다. 따라서 지도에 점이 있지만 얼마나 좋은지 표시하여 잠재고객이 한눈에 최고의 레스토랑을 결정할 수 있도록 합니다.

이 지도에 100개의 레스토랑이 있고 각각 아이콘 옆에 최대 5개의 별이 있다고 상상해 보십시오. 이는 많은 시각적 산만함이며 특히 주의를 기울이기 전의 순간에 사용자에게 압도적일 것입니다. 인터페이스에 그렇게 많은 시각적 정보를 집어넣으려고 하기보다는 순위를 지정하지 않거나 색상이나 대비를 사용하여 상대적인 점수를 표시하는 것이 더 나을 것입니다. <https://medium.com/towards-data-science/3-expert-data-visualization-tips-for-grabbing-readers-attention-206d8c4621bf>

“Plotly는 우리 비즈니스 개발 플랫폼의 중심에 있습니다… 우리는 방대한 양의 데이터를 빠르게 이해 및 분석하고 그 결과를 사용하여 수백만 달러의 투자 결정을 내릴 수 있습니다.”

— Jenya Kirshtein 박사 , C12 Energy의 과학 소프트웨어 엔지니어링

대시보드를 제대로 만든다는 것은 장식용으로만 사용되는 불필요한 대시보드 추가 기능인 차트 정크를 최소화하는 것을 의미합니다. 실제로 정보를 표시하는 데 사용되는 잉크의 비율인 데이터:잉크 비율을 최대화하려고 합니다.

데이터: 잉크 비율을 최대화하기 위해 Plotly의 기본 차트는 가는 회색 그리드 선을 사용하고 플롯을 둘러싸는 선을 사용하지 않습니다. 여기서는 Python, Pandas 및 Plotly 를 사용하여 주가 를 집계했습니다. 필요한 데이터만 한 곳에 모았습니다.

<https://medium.com/plotly/eight-helpful-tips-you-should-hear-from-visualization-experts-b3f289dacdeb>

보고서와 블로그 기사의 그래프 이후 사람들은 이야기에 더 관심을 갖게 되었습니다. 그래프에는 숫자, 날짜, 이름 등 다양한 데이터가 포함되어 있습니다.

이것은 모든 데이터를 한 번에 한 곳에 두어야 한다는 의미는 아닙니다. 불필요한 정보로 스키마를 로드하지 마십시오. 작게 두는 것이 좋지만 데이터는 확인되고 이해할 수 있습니다.

<https://medium.com/ux-planet/12-tips-to-design-better-data-visualization-c7740a4da5e1>

접근 가능한 세부 수준 표시 지나치게 세부적인 수준으로 차트를 채우지 않는 것이 중요 합니다. 예를 들어 일부 데이터 차트에는 너무 많은 레이블이 포함되어 있어 그 아래에 있는 그래프가 있어야 할 중요성을 잃습니다. 적절한 양의 정보를 차트에 넣고 많은 양의 데이터를 영향을 잃지 않고 작은 공간에 표시하는 창의적인 방법이 있습니다. 선이 시작되고 끝나는 위치, 격자선이 잘 배치된 요소 등을 고려하십시오.

막대형 또는 원형 차트의 영역과 산점도의 거품 크기는 모두 정보를 보다 조밀하게 압축하는 방법입니다. 예를 들어 시간 경과에 따른 세계 인구의 진행 상황에 대한 Hans Rosling의 유명한 산점도 는 정보 밀도가 매우 높고 시각적 스토리텔링의 좋은 예입니다. 애니메이션 차트를 만들 수 있는 도구가 없더라도 여기에서 작동하는 원칙을 정적 데이터 차트에 복제할 수 있습니다.

<https://medium.com/@hugocheyne/best-practices-for-building-informative-data-charts-f67158f9854c>

# 강력한 제목과 주석 설정

차트에 메시지를 명확하게 구성하는 강력한 제목을 지정하십시오. 훌륭한 제목은 차트를 더 기억에 남고 이해하기 쉽게 만듭니다.

필요할 때 주의를 끌기 위해 주석 달기 경우에 따라 일부 특정 데이터 요소에 주의를 기울이거나 차트에 추가 정보를 추가할 수 있습니다. 이를 위해 플롯에 텍스트 주석을 추가하여 정보를 확장할 수 있습니다. 얼마 전에 만든 차트의 예는 다음과 같습니다.

작성자의 이미지 이 예에서는 텍스트 주석을 사용하여 흥미로운 데이터 포인트(이상값)에 레이블을 지정하고 다른 방법으로는 식별할 수 없는 플롯에 추가 정보(영화의 25%가 손실)를 추가합니다.

그러나 텍스트 주석이 항상 유용한 것은 아닙니다. 주석이 필요한 항목이 없으면 차트에 강제로 추가할 필요가 없습니다. 또한 주석이 너무 많으면 차트가 복잡하게 느껴지고 가독성이 떨어질 수 있다는 점을 항상 염두에 두어야 합니다 . 최선의 판단을 통해 주석이 필요한 시기와 양을 결정하십시오.

플롯에 직접 레이블 추가 전통적으로 범례는 플롯의 상단이나 측면에 추가되었습니다. 그러나 레이블을 플롯에 직접 추가하면 플롯을 훨씬 쉽게 읽을 수 있습니다. 효과를 확실히 나타내려면 가능하면 선의 색상을 해당 레이블과 일치시켜야 합니다.

작성자의 이미지 더 좋아 보일 뿐만 아니라 플롯이 차지할 수 있는 공간도 증가하여 읽기가 더 쉬워집니다. 대안은 많은 공백이 있는 플롯 내에 범례를 배치하는 것입니다. 위의 예에서 왼쪽 상단은 완벽한 지점입니다.

물론 플롯 유형에 따라 이것이 어려울 수 있지만 가능할 때마다 일반적으로 읽기가 더 쉽습니다.

# 차트의 적절한 크기

차트의 크기를 적절하게 조정합니다. 각 축에 사용되는 눈금은 간격이 동일해야 합니다. 이는 차트가 올바르게 표시되는지 확인하는 빠른 방법입니다.

# 읽기 쉬운 글꼴과 크기

읽기 쉬운 제목, 축 레이블 및 범례의 글꼴을 선택하십시오. 사람들이 귀하의 메시지와 빠르게 연결되기를 원합니다. 이전 데이터 시각화 팁에서와 마찬가지로 라벨링과 관련하여 적을수록 좋습니다. 비슷한 이유로 여백을 남기고 더 적은 요소로 시각적 혼란을 줄이는 것 이상입니다. 항상 레이블을 그래픽으로 설명해야 하는 “가지고 있으면 좋은 것”으로 생각해야 합니다. 절대적으로 필요한 경우를 제외하고 레이블이 없다고 가정합니다.

나에게 완벽한 차트는 몇 가지 모양 외에는 아무것도 없습니다. 모든 것은 상대적입니다. 실제 값이 없더라도 한 항목이 다른 항목보다 얼마나 큰지 알 수 있습니다. 이제 당신의 상사와 데이터 담당자는 이 철학에 대해 격렬한 논쟁을 벌일 것입니다. 그들은 숫자가 없는 차트는 완전한 시간 낭비이며 완전히 터무니없다고 말할 것입니다. 좋아요, 그럼, 당신은 그들에게 물어볼 것입니다. 무엇이 라벨링되어야 합니까? (기본적으로 아무 것도 포함되어 있지 않습니다. 이제 무엇이 포함되어야 하는지 확신시켜 주십시오.)

이 시점에서 그들은 모든 것에 레이블을 지정해야 하고 축 선에는 20개의 눈금 표시가 필요하며 배경에 그리드가 있어야 차트 내의 모든 점이 어디에 있는지 볼 수 있으므로 모든 항목에 대한 실제 값을 식별할 수 있다고 말합니다. 10% 정확도 내에서 하나.

호그워시. 모든 것이 중요하다면 아무것도 중요하지 않습니다.

반드시 라벨을 부착해야 하는 부분에만 라벨을 부착하십시오. 중요한 차트의 4개 지점에 레이블을 지정합니다. 축은 아마도 각각 두 개의 눈금 표시(0 및 최대값)가 필요할 것이고, 축은 다른 레이블처럼 굵은 검정색이 아닌 밝은 회색일 수 있으며 데이터 뒤의 격자선을 제거할 수 있습니다.

사실, 이 차트의 전체 목적인 하나의 데이터 포인트를 나타내는 하나의 점을 제외하고 모든 레이블은 밝은 회색이고 더 작을 수 있습니다. (KWYRWTS를 기억하십니까? 이 목록의 첫 번째 데이터 시각화 팁은 무엇에 레이블이 지정되고 시각적으로 강조되는지 안내하는 데 도움이 됩니다.) 아, 그리고 우리가 하는 동안 구석에 있는 키를 죽이고 인라인을 사용합시다. 차트의 색상과 모양을 설명하는 레이블 지정.

주의를 기울이지 않는 잠재고객은 이제 판단할 항목이 줄어들고 귀중한 1/10초 동안 볼 올바른 것을 찾을 가능성이 높아집니다. .

다음 프로젝트를 1/10초 디자인 기술을 테스트할 수 있는 기회로 생각하십시오. 그리고 자신의 차트를 빠르게 살펴보세요. 보시고 미리미리 판단해 보세요. 어렵지만 한눈에 보는 청중과 심층 연구 청중을 고려하여 디자인하는 것이 매우 효과적입니다. 모든 사람이 전자이고 일부만이 후자라는 사실을 잊지 마십시오. <https://medium.com/towards-data-science/3-expert-data-visualization-tips-for-grabbing-readers-attention-206d8c4621bf>

# 출처 표기

투명성을 위해 항상 출처를 밝히십시오. 이것은 귀하에게 신뢰성을 제공하고 신뢰를 구축하며 시청자에게 더 많은 정보를 위해 소스를 방문할 기회를 제공합니다.

# 데이터의 논리적 구성

데이터를 논리적으로 구성합니다. 열과 막대를 값별로 정렬하여 한 눈에 쉽게 비교할 수 있습니다. 비교를 위해 원형 차트를 사용하지 마십시오. 원형 차트는 한눈에 비교하기 어려우므로 막대형 또는 세로 막대형 차트를 사용하는 것이 가장 좋습니다. 데이터를 시각화할 때 항상 생각하고 싶은 한 가지는 얼마나 많은 차원의 데이터를 표시할지입니다. 대부분의 그래프는 일반적으로 포인트를 얻기 위해 1-2 차원의 데이터만 필요합니다. 예를 들면 다음과 같습니다.

산점도에서 농구 선수의 키 x 몸무게 막대 그래프에서 로스앤젤레스 레이커스의 선수 높이 차원 수를 늘리면 단일 그래프가 더 복잡해져서 요점을 파악하기가 더 어려워집니다. 예를 들어:

산점도에서 농구 선수의 키 x 몸무게 , 30개 팀마다 다른 색상 점 . 다른 팀을 나타내는 30개의 개별 색상과 모든 팀 이름을 나열하는 범례가 필요하기 때문에 읽기 어려울 것입니다. 패싯의 마법이 빛나는 곳입니다. 우리 자신을 하나의 그래프로 제한할 필요가 없다면 어떨까요? 많은 사람들이 이런 식으로 생각하는 이유에 대한 제 가설은 우리가 단일 그래프로 제한되는 Excel에서 데이터를 시각화하는 데 익숙하기 때문입니다. ggplot에서 우리는 이러한 사고 방식을 깨뜨릴 수 있으며, 그렇게 하려면 한 줄의 코드만 있으면 됩니다. 패싯을 사용하면 시각화에 최대 2개의 추가 차원을 쉽게 추가할 수 있습니다.

좋은 Y축 시작점 선택

매력적인 비주얼을 만들려면 적절한 y축 시작점과 끝점을 선택해야 합니다. 어떤 사람들은 모든 그래프가 0에서 시작해야 한다고 주장하지만 이 접근 방식은 너무 엄격하고 많은 차트의 가독성을 파괴할 수 있습니다.

안타깝게도 y축을 어디에서 시작해야 하는지에 대한 훌륭한 경험 법칙은 없습니다. 대부분의 경우 y축을 0에서 시작하면 그래프를 읽기 어려울 수 있습니다. 그러나 확대하면 데이터의 작은 변화가 실제보다 훨씬 크게 보일 수 있습니다. y축을 0 값에서 시작하는 것이 가장 좋은지 아니면 다른 값에서 시작하는지 결정하려면 모든 고유한 상황에서 장단점을 고려해야 합니다.

규모를 데이터에 비례하게 유지 그래프에 있는 것보다 작거나 크게 보이도록 데이터를 변경하거나 이동하지 마십시오. 그렇게 하면 데이터를 잘못 표현하고 잘못된 결과를 전달할 수 있습니다. 또한 분석가가 나머지 숫자를 고려하지 않고 데이터의 특정 측면을 강조하려고 할 때 종종 배포됩니다. 따라서 명확한 결과를 전달할 수 있도록 시각적 개체의 크기를 조정하세요 .

항상 기준선 사용 작은 차이가 더 크게 보이도록 기준선을 0보다 큰 숫자로 시작하고 싶을 수 있지만 궁극적으로 오해의 소지가 있으며 노력을 약화시킬 것입니다. 예를 들어 보는 사람의 눈은 0보다 높은 기준선에서 시작하여 막대의 다양한 높이를 비교하여 막대 차트를 처리하므로 뇌가 이러한 시각적 비교를 수행하는 방식을 조작하고 잠재적으로 잘못된 결론에 이르게 됩니다. 항상 0부터 시작하십시오.

싫어하는 건 생략하지마 목표가 명확해야 하지만 다른 제안을 하는 데이터가 희생되어서는 안 됩니다. 다시 말하지만, 그것은 당신의 노력을 약화시킬 것입니다. 요점을 설명하기 위해 데이터를 생략해야 하는 경우 적절하게 뒷받침할 수 없는 결론을 구부리고 있는 것일 수 있습니다. 더 큰 그림을 두려워하지 마십시오. 더 좋은 점 은 내러티브의 일부로 불일치를 사용하는 것 입니다. 예를 들어 결과가 왜곡된 경우 왜 이런 일이 발생했는지 조사하고 결과의 일부로 “이유”에 답하는 것을 목표로 합니다.

<https://medium.com/@hugocheyne/best-practices-for-building-informative-data-charts-f67158f9854c>

# 적절한 색상 사용

색상을 사용하여 차트의 특정 부분에 주의를 끌 수 있습니다. 밝은 색상은 시선을 빠르게 끌어 메시지를 더 빨리 전달하는 데 도움이 됩니다. 무지개 또는 혼합 색상 팔레트를 피하십시오. 그들은 예쁘지만 반드시 효과적이지는 않습니다. 전체 차트에 대해 하나의 색상을 선택하거나 팝 색상을 활용하는 것이 좋습니다. 색상은 데이터 시각화에서 두 가지 주요 용도로 사용됩니다.

시각화를 더욱 매력적으로 만듭니다. 데이터의 추가 차원을 나타냅니다. ggplots에 색상을 지정하는 방법은 여러 가지가 있지만 간단하게 하기 위해 이 섹션에서는 제가 개인적으로 가장 좋아하는 Viridis 팔레트 에 중점을 둡니다.

다채로움: 차이를 쉽게 볼 수 있도록 가능한 한 넓은 팔레트를 사용합니다. 지각적으로 균일함: 값의 범위 전체에서 일관되게 서로 가까운 값은 비슷한 색으로 나타나고 서로 멀리 있는 값은 더 다르게 나타나는 색을 가짐을 의미합니다. 색맹에 강함: 위의 속성은 회색조 인쇄뿐만 아니라 일반적인 형태의 색맹을 가진 사람들에게도 적용됩니다.

<https://medium.com/@Infogram/dos-and-donts-data-visualization-tips-before-and-after-f1d65a7b6402>

색상은 그래프를 보기 좋게 만들 뿐만 아니라 데이터 분석에 기여해야 합니다. 이 대화형 지도에서 색상 척도는 미국 각 도시의 인구를 보여줍니다. 범례를 눌러 트레이스를 끄고 켭니다. 이 맵에 대한 자세한 내용은 Python 문서 를 참조하세요 . 또한 R 및 RColorbrewer 를 Plotly와 함께 사용하여 스펙트럼 팔레트를 만들고 Python 및 colorlover 를 사용하여 컬러 스케일을 사용할 수 있습니다.

<https://medium.com/plotly/eight-helpful-tips-you-should-hear-from-visualization-experts-b3f289dacdeb>

여기서 가장 좋은 방법은 색상을 사용하여 조직을 만드는 것입니다. 커뮤니케이션에 사용되는 모든 디자인과 마찬가지로 우수한 데이터 가시성 구조는 표준 원칙을 사용하고 요약으로 사용해야 합니다.

하나의 색상을 사용하십시오. 연속 데이터를 표시하거나 대비/강조 표시와 비교하여 사용합니다.

단어와 색상을 일치시킵니다.

색상은 디자이너 툴킷에서 가장 강력한 도구 중 하나입니다. 그러나 깨야 할 규칙이 있습니다. 이메일 디자인에서 3개 이상의 서로 다른 글꼴을 사용할 수 없으므로 여기에서 색상을 과도하게 사용하지 마십시오. 데이터에 가장 적합한 5~6개 이상을 선택하십시오.

<https://medium.com/ux-planet/12-tips-to-design-better-data-visualization-c7740a4da5e1>

다양한 색상과 형태를 차트에 배포 하면 보는 사람이 우선 순위가 지정된 요소에 끌리게 됩니다. 다양한 길이와 너비, 데이터 포인트의 모양을 변경하여 구분(예: 삼각형과 원), 인클로저 및 방향 실험, 흐림 및 곡률로 모양 조작과 같은 방법을 사용하는 것을 고려하십시오.

뇌가 시각적 정보를 받아들이는 잠재의식 과정을 전 주의적 처리 라고 합니다. Interaction Design Foundation 에는 더 깊은 수준에서 과학에 뛰어들어 연구 결과를 돋보이게 할 수 있는 몇 가지 유용한 리소스가 있습니다 .

잉크 대 정보 비율 최적화 보는 사람 이 보는 모든 것이 즉시 유용하고 유익하도록 같은 이야기를 전달하는 데 더 적은 줄을 사용 하십시오. 예를 들어 차트를 채우기 위해 점과 해치를 추가하지 마십시오. 블록 색상(예: 일반 파란색) 또는 빈 채우기도 마찬가지로 효과적일 수 있습니다.

더 적은 잉크를 사용하여 사용자가 한 눈에 더 많은 정보를 캡처할 수 있도록 하는 것은 Kevin McGurgan의 이 데이터 차트 비교에 명확하게 표시됩니다. 그러나 일단 특정 비율(“중간” 데이터 잉크 비율과 “높은” 데이터 잉크 비율을 비교할 때 표시됨) 미만이 되면 데이터에 대한 지식이 부족한 소비자는 쉽게 혼란스러워할 수 있습니다.

<https://medium.com/@hugocheyne/best-practices-for-building-informative-data-charts-f67158f9854c>

<https://medium.com/@jordi-camps/my-10-rules-for-effective-data-visualization-a01ba1077a93>