

**1- a. Trace um esboço do caminho que ele percorreu analisando os dados para responder essas perguntas.**

O video é bem interessante e em todas as análises que Hans Rosling faz ele mostra que temos sempre que observar as mudanças sociais antes das mudanças econômicas para entender como os dados estão mudando. Durante o video para mostrar que não temos mais um mundo tão separada quanto antigamente primeiramente ele analisa dados sobre fertilidade e com dados bons e com taxas de erros pequenas e analisou junto a expectativa de vida e chegou a conclusão que antes países com indústrias tinham famílias pequenas e vidas longas e os em desenvolvimento tinham famílias grandes e vidas curtas mas fazendo uma análise dos dados da década de 60 até agora os anos 200 os dados se agruparam bastante e agora temos vidas longas e famílias pequenas e um mundo totalmente novo como o próprio descreve. Em seguida Hans faz uma comparação entre os dados dos EUA e do Vietnã. Em 1964 os EUA tinham famílias pequenas e de vida longa e o Vietnã com famílias grandes e vidas curtas. Observou-se nos dados que mesmo durante a guerra houve melhora na expectativa de vida em ambos os países, ele mostra a evolução dos dados com os anos e cita eventos como o planejamento familiar, a entrada na economia de mercado e desistência do planejamento comunista, todos acontecidos no Vietnã e mostra que hoje o Vietnã tem a mesma expectativa de vida e o mesmo tamanho de família dos EUA. Logo após Hans mostra uma pesquisa que mostra que existe uma forte relação entre a mortalidade infantil e o PIB mostrando a distribuição de renda das pessoas de todo o mundo e observa que não existe mais intervalo entre rico e pobre como antigamente mostrando que o conceito de países em desenvolvimento é duvidoso. Em seguida Hans mostra onde está dividida a renda entre os países e mostra vários casos de como a distribuição de renda é desigual internamente nos países e em uma visão geral mostra que nos anos 70 a curva de diferença entre ricos e pobre era grande e hoje em dia não é mais. Por fim Hans fala sobre a importância dos países com os seus dados.

**1- b. Anote os termos ou ideias que você não entende ainda**

Entendi todos os termos e ideias usados no video.

**2- Qual o panorama geral dos profissionais de TI no Brasil sem ainda se preocupar com seus salários, apenas com sua localização, função, expertise...?**

Para importar os dados foi usado o comando:

```
salarios.ti.regioes <- read.csv("~/Analise de Dados I/problema  
1/problema01/salarios-ti-regioes.csv")
```

Para ter uma visão geral dos dados foi usado o comando

```
> summary(salarios.ti.regioes)
```

Primeiramente foi feita a pergunta:

Como estão divididos os profissionais em TI pelas regiões do Brasil, temos o resultado:

```
> regioao <- table(salarios.ti.regioes$Regiao)
> regioao
```

Centro-oeste	Nordeste	Norte	Sudeste	Sul
19	34	2	92	13

Como se classificam os profissionais em relação a graduação:

```
> graduacao <- table(salarios.ti.regioes$Pos.Graduacao.ou.Certificacao)
> graduacao
```

False	True
82	80

Como se classificam em relação a iniciativa de onde trabalham;

```
> iniciativa <- table(salarios.ti.regioes$Iniciativa.Privada.ou.Concursado)
> iniciativa
```

Na	Concursado	Privada
8	36	118

Qual a relação entre o tipo de cargo e a iniciativa de trabalho?

```
> iniciativaCargo <-
table(salarios.ti.regioes$Iniciativa.Privada.ou.Concursado,salarios.ti.regioes$Pos.Graduacao.ou.
Certificacao)
> plot(iniciativaCargo)
```

	FALSE	TRUE
NA	3	5
Concursado	16	20
Iniciaiva Privada	63	55

Qual a relação entre a região de trabalho e a iniciativa?

```
> iniciativaCargo <-  
table(salarios.ti.regioes$Iniciativa.Privada.ou.Concursado,salarios.ti.regioes$Regiao)  
> iniciativaCargo
```

	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
NA	2	2	0	4	0
Concursado	11	9	2	8	6
Ini. Privada	6	23	0	80	7

Com estes dados já podemos traçar um panorama onde nota-se que no Brasil os profissionais de TI estão bem balanceados em relação à sua formação e se trabalham na iniciativa privada ou não e nota-se que a maioria dos profissionais trabalham no sudeste e dentre estes a grande maioria trabalha na iniciativa privada.

**3- Os salários estão bem distribuídos no intervalo que vai do menor salário para o maior salário? Ou será que os salários são muito concentrados em um intervalo pequeno de valores? Será que existem salários atípicos (outliers)?**

```
> summary(salarios.ti.regioes$Salario.Bruto)  
> boxplot(salarios.ti.regioes$Salario.Bruto)
```

Com os dados obtidos no boxplot e no summary temos que existem 4 outliers com 3 recebendo entre 15000 e 20000 e um recebendo mais de 40000, em relação ao resto dos dados temos que os salários estão agrupados entre 2281 e 5333 com uma mediana de 3583.

**4- A distribuição que você encontrou é a que você esperava sem olhar os dados? Se não é, você tem alguma suspeita da razão para a diferença?**

Quanto à média o resultado bate com o que eu esperava, mas eu esperava que salários tão altos não fossem tão atípicos; a suspeita de porque isto aconteceu acho que foi por conta do pouco volume de dados, na minha opinião se tivéssemos mais dados acho que a média ainda ficaria a mesma, mas teríamos mais valores com salários altos.

**5- É interessante ver a média (mean) salarial para cada região. Como resultado, apresente uma tabela com o nome da região e a sua média salarial ordenada da região com a maior média salarial para a menor. Se você tivesse usado a mediana (median) em vez da média como métrica de interesse, o resultado seria o mesmo? Analise a situação e justifique. Aproveite e salve esta tabela em um arquivo do tipo csv.**

Primeiramente o comando usado foi:

```
> sort(tapply(salarios.ti.regioes$Salario.Bruto, salarios.ti.regioes$Regiao, mean),
decreasing=TRUE)
```

Centro-oeste	Nordeste	Sudeste	Sul	Norte
5620.474	4565.706	3962.227	3657.538	2552.500

```
> sort(tapply(salarios.ti.regioes$Salario.Bruto, salarios.ti.regioes$Regiao, median),
decreasing=TRUE)
```

Centro-oeste	Sudeste	Sul	Nordeste	Norte
5000.00	3582.95	3500.00	2900.00	2552.50

Nota-se que o valor da media foi maior devido aos outlier's encontrados nos dados.

```
> salarioPorRegiaoUsandoMedia <- sort(tapply(salarios.ti.regioes$Salario.Bruto,
salarios.ti.regioes$Regiao, mean), decreasing=TRUE)
```

```
> write.csv(salarioPorRegiaoUsandoMedia, "~/Analise de Dados I/problema
1/problema01/salarioPorRegiaoUsandoMedia.csv")
```

```
> salarioPorRegiaoUsandoMediana <- sort(tapply(salarios.ti.regioes$Salario.Bruto,
salarios.ti.regioes$Regiao, median), decreasing=TRUE)
```

```
> write.csv(salarioPorRegiaoUsandoMediana, "~/Analise de Dados I/problema
1/problema01/salarioPorRegiaoUsandoMediana.csv")
```

## 6- Se um profissional de TI deseja ter um salario melhor, qual a(s) dica(s) para ele?

Utilizando os dados da questão anterior vemos que a região que paga o melhor salario é a centro-oeste, fazendo outras pesquisas tais como:

pegando apenas os dados dos profissionais dessa região temos:

```
> salarios.ti.centro.oeste <- subset(salarios.ti.regioes, salarios.ti.regioes$Regiao ==
"Centro-oeste")
```

Analisando esses dados temos que:

```
> salarioPorGraduacao <- sort(tapply(salarios.ti.centro.oeste$Salario.Bruto,  
salarios.ti.centro.oeste$Pos.Graduacao.ou.Certificacao, median), decreasing=TRUE)
```

```
> salarioPorGraduacao
```

TRUE	FALSE
5500	3995

```
> salarioPorIniciaiva <- sort(tapply(salarios.ti.regioes$Salario.Bruto,  
salarios.ti.regioes$Iniciativa.Privada.ou.Concursado, median), decreasing=TRUE)
```

```
> salarioPorIniciaiva
```

NA	Concursado	Iniciativa Privada
6600	5650	4000

Como os dados não informados na iniciativa foram poucos e analisando a mediana por iniciativa e por graduação a minha dica seria que se o profissional almeja um salario melhor ele deveria se graduar e ser aprovado em um concurso na região centro-oeste.